Copyright

by

Nicholas Ryan Meyerson

2014

The Dissertation Committee for Nicholas Ryan Meyerson certifies that this is the approved version of the following dissertation:

Evolutionary and Functional Analyses of Primate Genes Reveal Critical Host-Virus Interactions

Committee:

Sara L. Sawyer, Supervisor

Robert M. Krug

Jaquelin P. Dudley

James J. Bull

Lauren I.R. Ehrlich

Evolutionary and Functional Analyses of Primate Genes Reveal Critical Host-Virus Interactions

by

Nicholas Ryan Meyerson, B.S.Phy.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2014

Dedication

For my mother and father:

You gave me life

You taught me love

You let me be me

"I hope you find it through this endless wandering."

Acknowledgements

I am merely a reflection of my family, friends, mentors, and students. I have become who I am today by emulating those that I admire the most. Throughout life I have been gifted incredible relationships with truly inspiring people. The following section is about those people—the ones that *really* made all of this possible.

I would like to begin by extending my gratitude to Dr. Sara Sawyer. I wandered into your lab with very little knowledge of biology and absolutely zero experience with a pipette. I was incredibly fortunate to find a mentor that invested so much time in my scientific apprenticeship. Thank you so much for challenging me, aggressively promoting my career, and never giving up on me, even when I make outrageous statements about wanting to go to law school. I hold you in the highest regard as a professor, a scientist, a mentor, a colleague, and most importantly, a friend. It has been a great honor to be a part of your burgeoning laboratory, and I hope it continues to grow in Colorado.

Working towards a Ph.D. in molecular biology is a bit like being dragged through mud for six years straight. I could not have done it without my lab mates being there to pull me through the especially murky parts. Dr. Ann Demogines, you were there with me from the beginning and have always been my main scientific confidant. You always kept me on track with my wild ideas and taught me how to focus my efforts on the bench to maximize my productivity. It has been a privilege to see you become a mother, although you gave me far more baby information than a man of my age can handle. Thank you for

making lab a fun place where I could express my ideas and ask as many questions as I wanted… I know I tend to have a lot.

Susan Rozmiarek, our fearless lab manager, where would the Sawyer Lab be without you? It is scary to even imagine that scenario. Thank you for keeping everything so organized for us and being so hospitable during our laboratory parties. You know how jealous I am of your beautiful home and it is so very kind of you to willingly share it with all of us. I would just like to reiterate in writing that I fully intend to recruit you away from whatever lab you are working in at the time that I start my own laboratory. I sure hope you will accept my offer. Dr. Paul Rowley, the man with infinite enthusiasm that brings his twin babies to pubs and hurls their dirty diapers at my face, I thank you for all of our wonderful conversations about the nuclear pore and how mesmerizingly beautiful of a machine it is. I will always cherish our first year at Cold Spring Harbor when we lived it up 'Craigie' style. Dr. Dianne Lou, my fellow graduate student, you and I can relate on so many levels with our experiences during grad school. Thank you for teaching me so many experimental techniques, you really are a great mentor and you had a whole troop of successful undergraduates to show for that. To the junior members of the lab (Alex Stabell, Maryska Kaczmarek, and Dr. Zhao Shan), I hope that I have been as helpful in your training as those I have mentioned above. You all have the potential to be great scientists, and I cannot wait to see what you accomplish in the near future. Here's to Freddie Fridays, and to working overtime all the time, because deep in our hearts we know we love it.

I could not have been luckier when the TRIM25 project was dropped into my lap. I remember Sara calling me into her office and saying, "Well, get ready to become a flu biologist." Putting the project itself aside, the reason I was so lucky is because it was through this project that I forged a relationship with Drs. Robert Krug and Ligang Zhou. I was welcomed into the Krug Lab with open arms and infinite resources. It felt like my second home. Dr. Krug, thank you for keeping your door open for me. You did not only listen to my ideas, but you carefully considered them and made me feel like a colleague. We really hit a home run with TRIM25, and it would not have been remotely possible without you. Ligang, the secret weapon of the TRIM25 project, I thank you infinitely for your patience in teaching me how to culture and study influenza virus. Our project was so confusing for so long, and I know you share the excitement I feel at having figured out the mystery.

I would like to thank all of my remaining committee members: Drs. Jaquelin Dudley, James Bull, and Lauren Ehrlich. You all kept your doors open for me and were willing to discuss absolutely anything related to science or the graduate school experience. It was during my committee meeting this past April that I realized I had to stay in science. It was recommended that I give up my favorite project in order to graduate earlier. It is hard to explain how that made me feel, but I just knew I could not do it. It was then that I realized my obsession and that I would spend the rest of my life in science. Thank you all for providing the forum for me to grow professionally.

I had the unique opportunity to be a part of the Woods Hole course on Molecular Evolution for three years in a row. Thank you Dr. David Hillis for nominating me as a

teaching assistant for this course. It was a magical and reinvigorating experience and I hope I can return to this course in the future. I would also like to thank Dr. Conor Meehan, a fellow teaching assistant from this course, for quickly becoming one of my best friends. I am saddened that we do not see each other more and it is likely that we will not connect again at Woods Hole, but who knows, maybe one day we'll get that super lab we always talk about. It *will* be called Meyerson Lab, non-negotiable.

To all of the students I have had in classes and as tutees, you gave me more insight than you will ever know. You challenged me to obtain a master's knowledge of subject matter and always kept me on my toes with your questions and general concerns about academic life. I hope you all find the satisfaction in learning that I have, it is without a doubt the finest pleasure in life.

To all of my friends outside of science, especially my high school friends from Humble, you made this whole experience real. Perspective is everything in science, and you all provided me with exactly that. I will forever treasure all the late nights in the studios, playing music and celebrating our togetherness. I am so lucky to have had the same friends for so long, I will miss you all incredibly as I begin the next stage of my career in Colorado.

My mother and father have invested more in me than anyone else on this planet. They taught me about life and how to enjoy it, and in doing so projected me into a world where I see no boundaries. They always made me feel like I was in complete control of my future, and never judged my ambitions or pushed me away from things I was passionate about. I whole-heartedly thank you both for supporting me throughout my

endeavors. To my brother, sister, and all of my extended family, I have a tremendous amount of pride for where I came from. Thank you all for being invested in my education, your support is taken with gratitude.

And finally, I give my eternal appreciation to my fiancée, Lauren Zeni. You have supported and loved me unconditionally throughout this entire process, which I know was no easy task. I am so lucky to have found someone that is accepting of who I am and who I want to be. We share the same vision for our future, and I could not be more excited to marry you next year. You are my muse, my guiding light, and my true love.

Evolutionary and Functional Analyses of Primate Genes Reveal Critical Host-Virus Interactions

Nicholas Ryan Meyerson, Ph.D.

The University of Texas at Austin, 2014

Supervisor: Sara L. Sawyer

Viruses exert a tremendous evolutionary pressure on their hosts. By hijacking cellular machinery and resources, viruses have been wildly successful at infecting and propagating throughout all domains of life. In the following dissertation, the interactions between primates and some of the viruses that infect them are examined through an evolutionary lens. I begin by introducing the long-standing battle between mammals and viruses that has raged on for hundreds of millions of years. I propose a theoretical framework to understand how slowly evolving mammals are able to keep pace with rapidly evolving viruses, and how we might use this framework to monitor future virus outbreaks.

The core of my analyses stems from an evolutionary concept known as the host-virus arms race. This tug-of-war for survival between hosts and viruses leaves an imprint in the DNA of each organism involved that can be detected using statistical analyses. In

Chapter 2, I describe these analyses in great detail and perform many tests to ensure that they are being used and applied appropriately.

The remainder of my studies focuses on detecting novel signatures of positive selection in primate genes that are likely caused by ancient host-virus arms races. I characterize the evolutionary history of several primate genes that have been implicated in viral life cycles and provide functional evidence that viruses drove their rapid divergence. In doing so I make three important discoveries. First, I characterize a genetic variant of *CD4*, the cellular receptor for HIV-1, in an owl monkey species that could make them a viable HIV-1 model system. Second, I show that gorilla-specific mutations in *RANBP2*, a gatekeeper of the cell nucleus, can inhibit HIV-1 infection. And finally, evolutionary signatures in *TRIM25*, a component of the innate immune system, revealed its ability to inhibit influenza A virus replication by binding incoming viral ribonucleoproteins.

Table of Contents

List of Tables

List of Figures

Chapter 1*

Two-stepping through time: mammals and viruses

Recent studies have identified ancient virus genomes preserved as fossils within diverse animal genomes. These fossils have led to the revelation that a broad range of mammalian virus families are older and more ubiquitous than previously appreciated. Long-term interactions between viruses and their hosts often develop into genetic arms races, where both parties continually jockey for evolutionary dominance. It is difficult to imagine how mammalian hosts have kept pace in the evolutionary race against rapidly evolving viruses over large expanses of time, given their much slower evolutionary rates. However, recent data has begun to reveal the evolutionary strategy of slowly-evolving hosts. We review these data, and suggest a modified arms race model where the evolutionary possibilities of viruses are relatively constrained. Such a model could allow more accurate forecasting of virus evolution.

* Dr. Sara L. Sawyer helped to formulate the ideas put forth in this chapter. This work was published in Trends in Microbiology [127]. Permission to adapt the contents of the publication was acquired from Dr. Sawyer and the publishing company.

The microscopic and sub-microscopic parasites can evolve so much more rapidly than their hosts that the latter have little chance of evolving complete immunity to them… The most that the average species can achieve is to dodge its minute enemies by constantly producing new genotypes. -- J.B.S. Haldane, 1949 (1)

Ancient relationships between mammals and their viruses

Recent studies have unearthed a treasure trove of prehistoric virus 'fossils,' viral genomes or genome segments frozen millions of years ago as integrated copies in the genomes of diverse animal hosts (see (2) and references therein). The fact that these integrated viral fossils can be easily recognized as belonging to modern virus families is stunning, given the fact that modern exogenous viruses have replicated and evolved for many millions of years since these viral fossils were captured (2). Despite high rates of mutation, the evolution of virus sequence is clearly constrained. This constraint comes partially from intrinsic selective forces that limit virus evolution, such as selection for modulation of pathogenicity to the host, and the structural constraints of the virus itself. Other major constraints on virus evolution come from the diverse immune strategies imposed by hosts. Cumulatively, these constraints act together to limit all aspects of virus evolution, from the swarm of variants produced in a single host to the evolution of expanded host range. These newly identified fossils indicate that constraint on virus evolution may be far greater than has previously been appreciated.

The discovery of these fossils also confirms that mammals have evolved to their

current form within a landscape full of diverse viral threats, a situation that has dramatically influenced their own evolution. Although constrained in many ways, viruses still evolve much more quickly than the hosts that they infect. The response time for evolutionary adaptation by viruses can be nearly instantaneous, whereas mammals reproduce on the scale of years or decades. The dominant force enabling mammals to counteract the extreme genetic diversity of the viruses that they face is generally thought to be the adaptive immune system. Adaptive immunity employs gene rearrangements performed during the lifetime of an individual host, creating a nearly infinite spectrum of receptors and antibodies that evens the playing field between the host and rapidly evolving viral pathogens. But the first line of defense in fighting infection, and one that is thought to be successful against the vast majority of pathogens encountered, is the hard-wired innate immune system. Innate immunity is executed by genes that must function strictly in the form in which they were inherited. The human genome has approximately 1,000 genes dedicated to defense or immunity, most of which cannot diversify in the course of a single host lifetime (3). Somehow all of these genes must remain honed against their viral targets even though they are trapped in a slowly evolving mammalian genome (4). In recent years, molecular evidence has emerged, largely from the HIV field, that describes how hosts respond to infection over evolutionary time. Although the strategy of viruses is to rapidly adapt to new challenges, the strategy of hosts is, by necessity, altogether different.

Host immune complexity dramatically limits the escape options available to viruses

Genes involved in immunity and defense become effective against their viral targets by natural selection over many generations, with individuals encoding less effective alleles dying from infection with bias (5). Once effective immunity alleles become common, viruses are expected to counter-evolve, thereby placing selective pressure back on the host species. These dynamics create an ever-escalating genetic 'arms race' between host and virus that results in the rapid evolution of both (6). Genetic arms races have been shown to play out predominantly through host and virus proteins that interact directly. For instance, host major histocompatibility complex (MHC) genes encode receptors that present peptides at the cell surface for recognition by T cells. MHC class I genes have been acutely selected for alleles that effectively present viral peptides and, as a result, MHC genes are highly divergent between and within species (4). Viruses, in turn, evolve under selection for mutations that prohibit presentation of viral peptides by these receptors. In the past ten years, an entire landscape of membrane-bound or cytosolic viral sensors have been discovered to act in innate immunity (7-10). These include a large number of constitutively expressed proteins called 'restriction factors' that recognize viruses and inhibit their replication directly (11). In primates, these proteins include tetherin/BST-2 and members of the APOBEC3 and tri-partite motif (TRIM) families (12-14). For example, the TRIM5α protein interacts with the capsid core of retroviruses as they enter the cytoplasm of an infected cell (Figure 1-1A). Different primate orthologs of TRIM5α have recognition specificity for different retroviral capsids, and infection is only blocked when interaction occurs (15). Physical interactions, such as

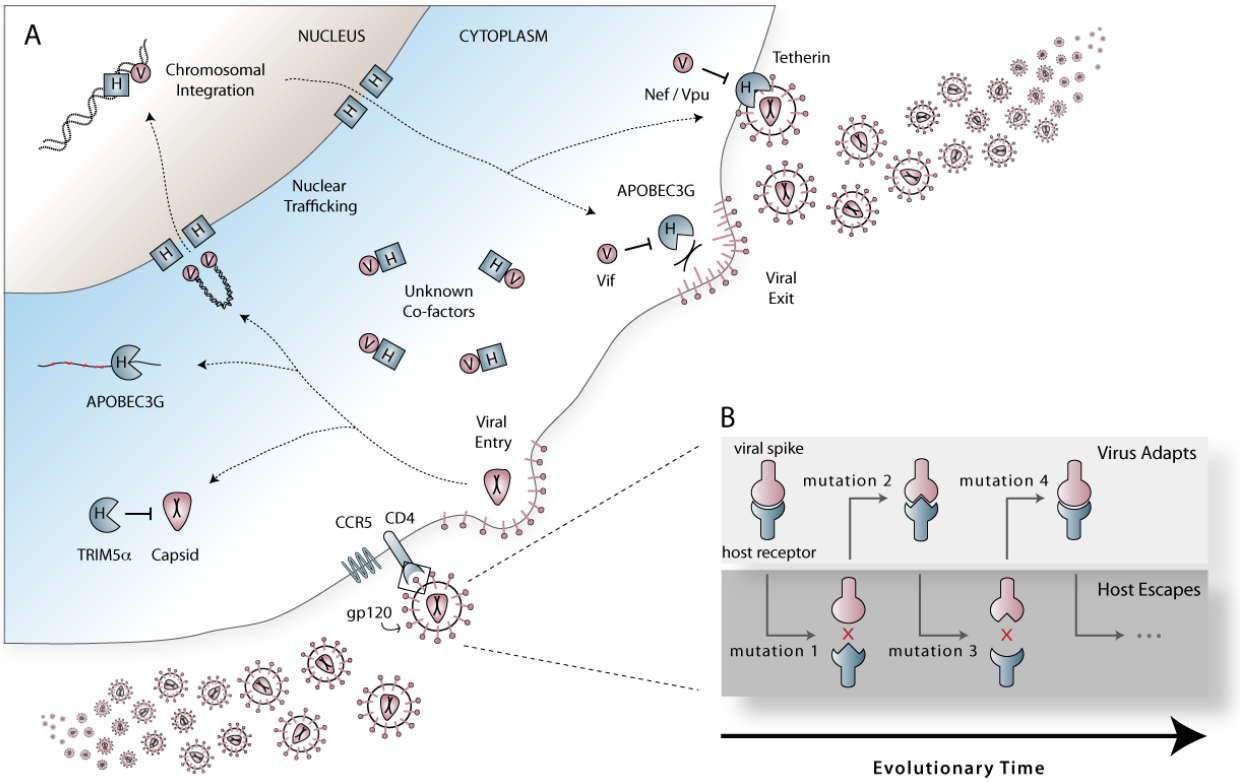


Figure 1-1: Evolutionary arms races involve interacting host and virus proteins. (A) A schematic of the HIV/SIV lifecycle is shown, illustrating some of the instances where virus proteins (V, red) are known to interact with host proteins (H, blue). In some cases these interactions involve host cofactors (squares and cell surface receptors) that are hijacked by the virus for replication. Hosts also encode antiviral restriction factors (pie shapes) that inhibit viral replication through various mechanisms. For instance, TRIM5α interferes with capsid cores, APOBEC3G hypermutates viral genomes, and tetherin inhibits virus budding. **(B)** The cellular receptor, CD4, is used to illustrate a hypothetical arms race scenario. Interaction of the viral spike glycoprotein with CD4 results in virus entry into the cell. Over time, any mutation in CD4 that reduces the strength of this interaction will be preferred by natural selection acting on the host population. Selective pressure will then be placed on the virus population for a mutation in the glycoprotein that re-establishes interaction with CD4. This back-and-forth interplay will result in the rapid fixation of mutations that alter both protein sequences.

the one between TRIM5α and capsid, are fertile ground for arms race dynamics (16, 17). In many cases, viruses also encode proteins that antagonize host immunity pathways (10, 13), and these physical interactions can also be subject to arms races (18-20).

Genetic arms races are fundamentally dialogues of call and response between host and viral genomes. However, since arms races unfold over evolutionary time scales, how do hosts actually compete, given the dramatically different evolutionary rates that usually divide mammalian hosts from their viral pathogens? The answer to this conundrum lies largely in the complex and multi-faceted nature of the immune system. Novel mutations in viral genomes that allow escape from one arm of the immune system will only be viable if they do not make the virus susceptible to other immune strategies. Viral evolution is further limited by interactions with proviral host proteins. As obligate parasites, viruses survive by hijacking host proteins (called cofactors) for processes such as cellular entry and nuclear trafficking (Figure 1-1A). For instance, cellular entry of HIV requires the human cell surface receptor CD4. A co-receptor is also required for entry, which for most HIV strains is the chemokine receptor CCR5 (Figure 1-1A)(21). In large-scale screens, HIV and influenza have each been shown to require several hundred cofactors for replication in humans cells (see (22, 23) and references therein). It should be noted that allelic versions of cofactors that lack compatibility with viruses are just as effective at blocking infection as potent immunity alleles, and possibly more so. For example, some humans encode a variant allele of *CCR5*, *CCR5Δ32*, where a 32 base pair deletion gives rise to a defective receptor that is not expressed on the cell surface (21). Those individuals homozygous for this allele are almost completely resistant to HIV

infection, and even heterozygous genotypes afford some protection due to reduced expression levels of CCR5. Why this allele exists is unknown, but it is highly relevant in a population that is now infected with HIV. In this respect, cofactors can also be engaged in arms race dynamics with viruses (24-27). In such cases, the host genome experiences selection to encode cofactors that are non-optimal for viruses, and viruses are selected to efficiently utilize available host cofactors (Figure 1-1B). New viral variants will only be viable if they retain all necessary cofactor interactions, and simultaneously avoid fatal interactions with immune system proteins.

Selective pressures exerted by the many host immunity proteins and cofactors in combination funnel viruses into a very small mutational space for escape and adaptation. This is supported by many examples where viruses repeatedly escape through the same amino acid change, not just through mutation from one amino acid to any other. For example, in experimental cross-species infections of simian immunodeficiency virus (SIV) from sooty mangabeys (SIVsm) into rhesus macaques, viral replication was initially weak due to restrictive *TRIM5* alleles (28). However, viral escape occurred in four different macaques and, in all cases, involved the same single amino acid change (R97S) in the capsid protein. In another example, there have been multiple independent cross-species transmissions of SIV from chimpanzees (SIVcpz) and possibly gorillas (SIVgor) to humans, giving rise to HIV-1 groups M, N and O. A single amino acid position in the *gag*-encoded matrix protein underwent the same mutational substitution (M30R) in all of these cross-species transmissions (29). When HIV-1 was passaged back through chimpanzees, this mutation reverted. Like these two examples from the HIV

field, there are many amino acid positions in the influenza genome that have characteristic mutations depending on the source species, suggesting that the web of selective constraints in each species repeatedly results in viruses acquiring signature adaptive mutations (30). The focused routes of adaptation seen in these examples are a reflection of the massive and multi-factorial constraints imposed by the host. This serves as an important reminder that viral evolution studies are most meaningful when conducted in animal models, as a greater range of viral mutations can be explored in cell culture assays than will actually be viable in the infected host.

Arms races play out in populations of hosts, which slows the rate of viral adaptation

Each step in an evolutionary arms race begins with selection for an advantageous mutation in a population, either host or viral (Figure 1-2). Only when adaptive mutations go to fixation in either a host or viral population is the arms race permanently moved onward. This is probably a relatively rare event, and the fixation of a potent mutation in a host or viral gene could theoretically end an arms race forever, driving the other party to extinction. Instead, the Red Queen hypothesis (31) predicts that selective pressures exerted by hosts and viruses upon one another will often result in oscillating allele frequencies in both populations. This is because selective pressure for viruses to counter-evolve will not be strong until a significant number of potential hosts are of a resistant genotype. This provides an important advantage to the host, because it slows the spread of novel, adaptive viral variants through host populations.

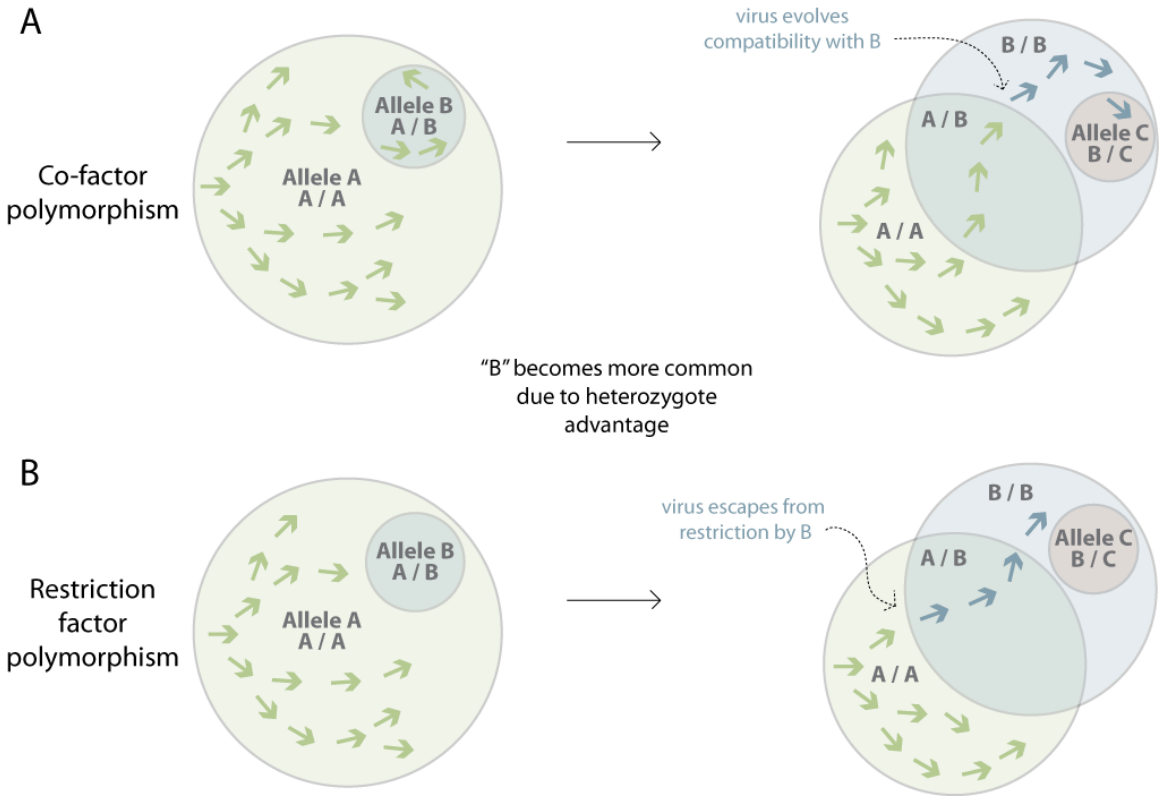


Figure 1-2: Arms races play out in populations.

Figure 1-2: Arms races play out in populations. The circles represent populations of hosts, with the circle sizes representing the relative frequencies of different alleles of a single host gene. The spread of a virus (arrows) through this population of hosts is illustrated. Overlapping regions represent heterozygous hosts and non-overlapping regions represent homozygous hosts. **(A)** In this schematic, the effect of polymorphism in a cofactor gene is illustrated, where the effect on susceptibility would be predicted to be semi-dominant (similar to the *CCR5Δ32* mutation discussed in the text). Allele A encodes a cofactor that can be exploited by the virus, whereas allele B encodes a cofactor variant that is resistant to viral exploitation. The virus will be able to replicate in both A/A and A/B individuals, although presumably less well in A/B heterozygotes. Eventually, allele frequencies in the host population will shift due to the partial protection afforded by the A/B genotype. As allele B becomes more common, there will be selective pressure on the virus to better utilize the B cofactor (blue arrows). A new resistant allele, C, might also emerge or pre-exist in the host population. **(B)** In this schematic, the effect of polymorphism in a restriction factor gene is illustrated, where allelic versions that restrict a virus are predicted to have a dominant influence on susceptibility to infection. The major allele, A, encodes a restriction factor that is inactive against a circulating virus population. Allele B encodes a restriction factor that is effective against this virus. When allele B is rare, all carriers are assumed to be heterozygous. Because A/B individuals are protected from infection, allele frequencies shift such that allele B is now more common, giving rise to B/B homozygotes. The diminished reservoir of A/A homozygotes places pressure on the virus to 'escape' (blue arrows) restriction by the B restriction factor. A new allele with potency against the virus, C, might also emerge or pre-exist in the host population.

The case of CCR5, the HIV cellular co-receptor mentioned above (Figure 1-1A), illustrates this concept. Most SIV strains also use CCR5 as a co-receptor, but some sooty mangabeys and red-capped mangabeys encode defective alleles of *CCR5* (32, 33). Despite the fact that these alleles are relatively common, the SIV strains that infect these two species (SIVsmm and SIVrcm) replicate perfectly well, even in individuals that are homozygous for defective *CCR5*. This is because in both cases these viruses have evolved to use additional co-receptors. This is the type of escape that can be expected in the face of common resistance alleles. In humans, the defective *CCR5Δ32* allele is quite common in some parts of the world, with about 18% of Caucasian individuals being heterozygous and 1% being homozygous (21). Similar to sooty and red-capped mangabeys, a few *CCR5Δ32* homozygous humans have also been reported to be infected with HIV, again through mutations of the HIV surface protein that allow use of an additional co-receptor (CXCR4 in this case) (21). This viral escape through co-receptor switching also happens in many late-stage HIV-1 patients who are wild-type for CCR5, and in patients treated with the CCR5-inhibiting drug Maraviroc (34). In both cases this is presumably because the preferred target cells become more scarce or unavailable. Due to the HIV/AIDS epidemic, there continues to be a rise in the frequency of hypomorphic alleles of *CCR5* in Africa (35), suggesting that HIV might continue to evolve the ability to use new co-receptors as highly susceptible hosts diminish.

The scenario just described is but one part of a larger picture. In fact, probably hundreds of host loci additively contribute to viral susceptibility. There are many human genes, acting in diverse cellular pathways, in which genetically-encoded polymorphic

variants can influence the outcome of viral exposure or infection (36, 37). Primates encode entire families of several key proteins involved in virus recognition, such as the APOBEC3, TRIM, MHC, killer cell immunoglobulin-like receptor (KIR), and interferon-inducible transmembrane (IFITM) protein families (9, 14, 38, 39). Further, each of these loci might have many co-circulating alleles in a population; the human *HLA-B* locus of the MHC has over 800 reported alleles (7). Diploid hosts can carry two different alleles at each locus, each with distinct viral targets. For dominant immunity factors, this means twice the specificity and therefore increased resistance to infection. Heterozygote advantage has been directly demonstrated at MHC loci in both humans and macaques (40, 41). Different alleles of *TRIM5* found in primate populations have also been shown to have different viral specificities, possibly leading to a scenario where balancing selection operates to maintain multiple alleles in populations (28, 42). With all of these considerations, it is unlikely that any two individuals in any mammalian population have the exact same genetic immunity profile. A key point is that, when viruses evolve to escape resistant host genotypes in the context of a host population, this will provide them access to only a limited number of new hosts due to the mosaic of host genotypes involved. The broad genetic diversity between immune system components of different individuals should dramatically slow the spread of viral escape variants through host populations, and is yet another weapon in the arsenal of the slowly evolving host.

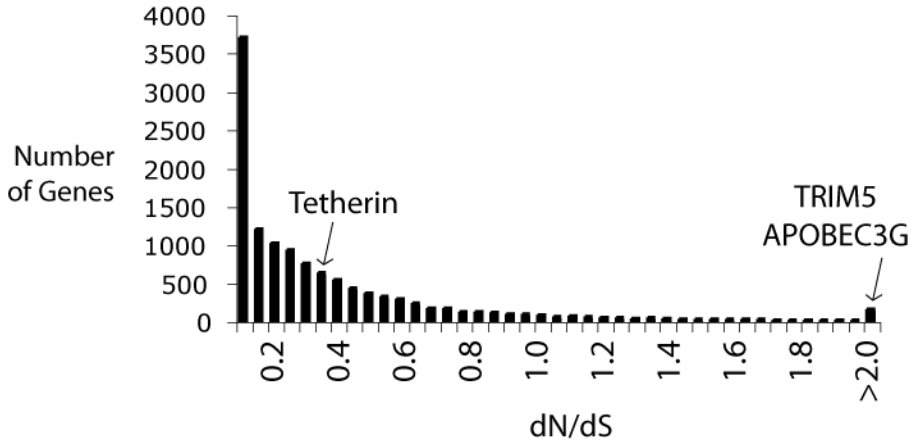


Figure 1-3: Primate restriction factor genes are rapidly evolving. The distribution shows dN/dS values previously determined for 13,454 human-chimpanzee orthologous gene pairs (43). Adaptive, gain-of-function mutations commonly arise from point mutations that change an amino acid in the encoded protein. When genes are experiencing sequential rounds of positive selection for new adaptations, as in the arms race scenario, they will retain a higher proportion of non-synonymous mutations (dN) than synonymous mutations (dS) in the domains critical for governing the physical interaction. Domains under positive selection will thus accumulate a characteristic signature of dN/dS > 1. dN/dS values for three primate retroviral restriction factor genes are indicated on this distribution. As expected, restriction factor genes such as *TRIM5* and *APOBEC3G* have some of the highest dN/dS values in the human genome (17, 38). Although it is known that codons within *tetherin* are evolving under positive selection (18-20), a full length gene analysis reveals a much lower dN/dS value (dN/dS ~ 0.3). *Tetherin* is either under less intense selective pressure for adaptation than *APOBEC3G* and *TRIM5*, or is more constrained by its other cellular roles. Both tetherin and TRIM5α are known to function in host roles other than retroviral restriction (44-46). Additional evolutionary constraint comes from the other host proteins with which restriction factors must interact to execute restriction (13, 47).

Single point mutations can lead to major adaptations during evolutionary battles

Because host genes engaged in evolutionary arms races are under strong selective pressure to change and adapt, they often evolve at a faster rate than other genes. In fact, immunity genes are some of the most rapidly evolving mammalian genes, acquiring unexpectedly high numbers of non-synonymous nucleotide substitutions (3, 48, 49). The

genes of all three of the retroviral restriction factors diagrammed in Figure 1-1 have accumulated such mutations more rapidly than the bulk of human genes (Figure 1-3). Importantly, this pattern indicates that single amino acid substitutions in a protein can be adaptive in the context of an arms race (17, 50, 51). This is critical to the success of the host because simple point mutations are the most common and abundant form of genetic variation upon which natural selection can act.

The power of point mutations can be well demonstrated with examples from the functional characterization of these same three retroviral restriction factors. Single amino acid changes in TRIM5α can be highly adaptive, as some allow recognition of different mammalian retroviral capsid types (52-56). Conversely, single amino acid changes in retroviral capsid proteins can also be adaptive by decreasing susceptibility to TRIM5α (28, 57-60). Another restriction factor, tetherin, is a cell surface, membrane-bound protein that prohibits budding of retroviruses as well as filoviruses, herpesviruses, and arenaviruses (Figure 1-1A)(13). Viruses encode countermeasures to tetherin, including the SIV antagonist Nef and the HIV-1 antagonist Vpu, and single point mutations in tetherin can modulate sensitivity to these viral antagonists (18-20). The APOBEC3G protein is a restriction factor with activity against a broad range of viruses, but is neutralized by the HIV/SIV accessory protein Vif (Figure 1-1A)(47). A single amino acid change in APOBEC3G can make it insensitive to Vif (61-63), and single amino acid changes in Vif can alter specificity for APOBEC3G (64). Thus, exquisitely small biochemical changes, at the level of a single amino acid, can enhance or reduce affinity between players in an evolutionary arms race.

In some cases, changes more dramatic than point mutations occur in the context of arms races, and these often end one arms race and start another. For instance, in two different primate lineages *TRIM5* has fused with the gene encoding cyclophilin A to produce a TRIM-Cyp restriction factor with a novel retroviral recognition domain, significantly changing the terms of the arms race between TRIM5α and capsid in those species (16). In humans, tetherin has acquired a deletion in the binding site for its historical SIV antagonist, Nef, freeing tetherin from that arms race until HIV-1 evolved a novel way to neutralize human tetherin using the viral protein Vpu (19, 20, 65). Adaptive point mutations in the ligand-binding surface of primate CCR5 that influence HIV and SIV binding are not common. Instead, as previously discussed, *CCR5* seems to fight the arms race through protein downregulation at the cell surface, with frame-shifted null alleles having arisen independently in humans and two additional SIV-infected primate species (32, 33, 66). Although highly effective, major adaptive changes such as these are expected to occur less often than adaptive point mutations, simply because they involve rarer genetic events.

Rock-paper-scissors: are arms races cyclic?

Our studies and those by other groups reveal that host restriction factors may be resampling a small number of biochemical forms repeatedly as arms races play out over evolutionary time. For instance, a single amino acid mutation in human TRIM5α (R332P) can largely restore to this protein the ability to recognize and restrict HIV, and is therefore an important determinant of retroviral specificity (54, 55). Interestingly, three

amino acids, arginine (R), proline (P), and glutamine (Q), have been repeatedly sampled by natural selection at position 332 over tens of millions of years of primate evolution (Figure 1-4A) (28). The R-, P-, and Q-bearing forms of TRIM5α each target different suites of retroviruses (53-60). Amazingly, R, P, and Q are each encoded by different alleles co-circulating in sooty mangabey populations, whereas P and Q are each encoded by different alleles found in rhesus macaque populations (Figure 1-4A) (42, 67). This pattern might reflect ancient polymorphism that has survived all or some of simian primate speciation, rather than recurrent mutation to the same few amino acids (42). Long-term maintenance of such polymorphism could occur if there were balancing selection acting on these alleles (42, 67). However, polymorphism has rarely been shown to survive even a single speciation event. Regardless of how this striking pattern arose, continued selection for three specificities must be invoked to explain it.

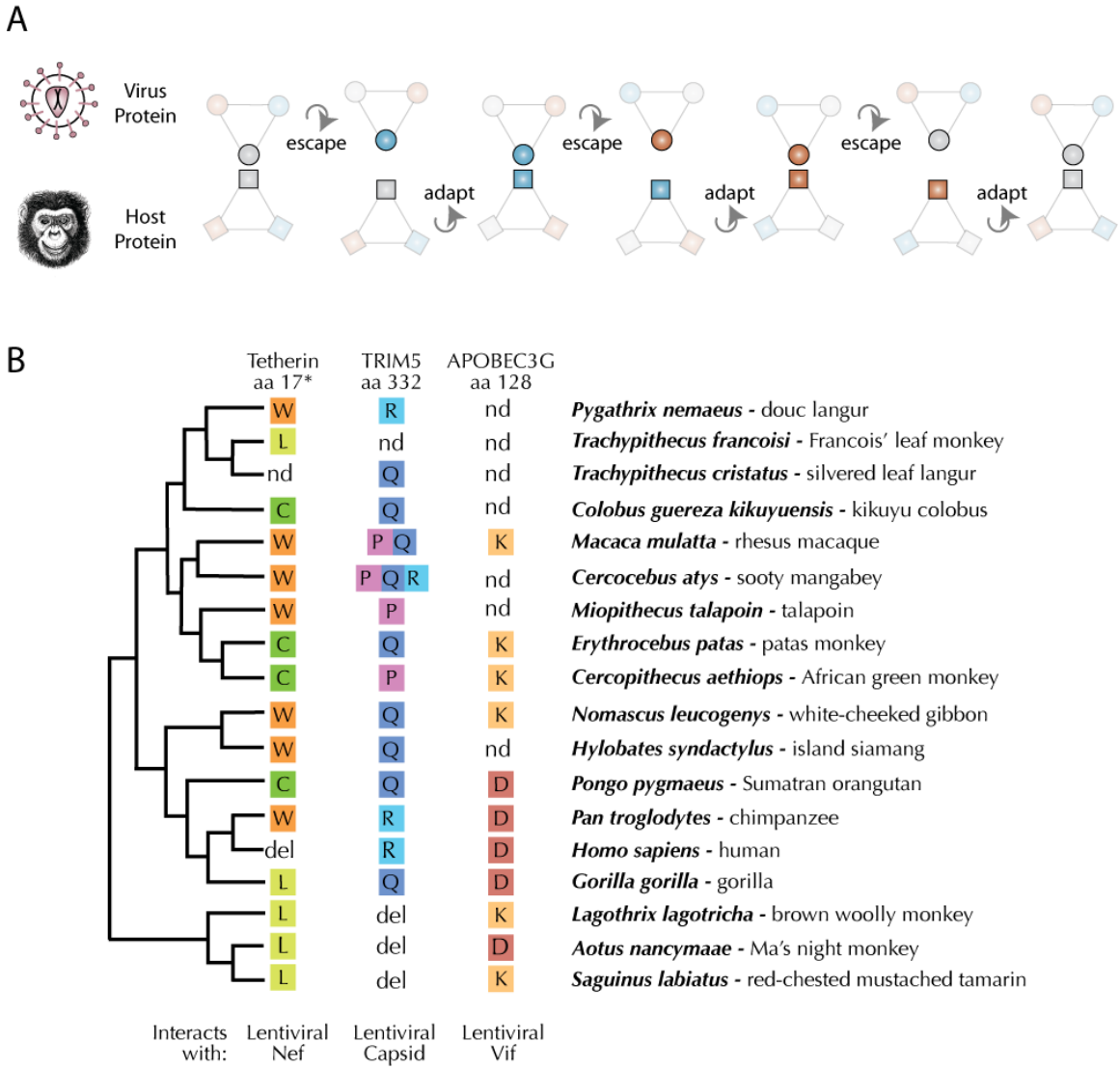


Figure 1-4: The cyclic arms race: a game of rock-paper-scissors?

Figure 1-4: The cyclic arms race: a game of rock-paper-scissors? (A) Molecular evidence for the rock-paper-scissors model. For each restriction factor, a key residue known to alter specificity for retroviral targets is illustrated, and the amino acid (aa) encoded at that position is reported for a panel of primate sequences available on Genbank. In two cases, SNPs have been noted at these positions, so multiple amino acids are listed. Amino acid coordinates refer to the human protein except in tetherin (*), where amino acid 17 is the chimpanzee coordinate, since amino acids 14-18 are deleted in the human sequence. Abbreviations: nd, not determined; del, deletion; W, tryptophan; L, leucine; C, cysteine; R, arginine; Q, glutamine; P, proline; K, lysine; D, aspartic acid. **(B)** The illustration shows the physical interaction between a host restriction factor protein (square) and a virus protein (circle). In the rock-paper-scissors model, both the host and the virus have only a small number of biochemical variants with which to compete (3 in this example), because of the many constraints placed on each. These variants are symbolized by different colors. When colors are matched, the host restriction factor successfully inhibits viral infection. However, escape by the virus is also possible. The host population will then respond through selection for an adaptive mutation that re-establishes interaction. The virus might re-sample previous states (gray circle) that will now be resistant in the context of the new host genotype (orange square).

It is unknown why TRIM5α has resampled three amino acids (R/P/Q) at position 332 over evolutionary time. Only three amino acids at this position would be necessary if there are just three amino acids possible at the cognate position(s) in capsid. If true, this would represent a highly constrained arms race, one that is perhaps consistent with the many constraints detailed above. The consistent cycling over evolutionary time between three amino acids (R/P/Q) is perhaps indicative of a rock-paper-scissors game where the choice of weapons on both sides is limited, and therefore must be recycled (Figure 1-4B). The evolutionary histories of other retroviral restriction factors, APOBEC3G and tetherin, also reveal resampling of a small set of amino acids at key positions that critically govern viral specificity (Figure 1-4A). This is interesting, given the vastly different viral repertoires and modes of action of each of these three restriction factors. Small insertions and deletions in the regions of these critical residues have also been observed (17, 19, 68), but these more dramatic mutations may sometimes change the interaction surface in a way that still fits into the cyclic pattern of recycled biochemical forms, or may introduce a new one.

In reality, physical interactions involve three-dimensional protein surfaces, not single amino acid sites, so changes at single sites must be considered in the context of the rest of the protein. This can be demonstrated with examples from TRIM5α biology, where R332 in the context of the human TRIM5α protein does not restrict HIV, but R332 in the context of sooty mangabey TRIM5α does (42, 54, 69). It is known that other amino acids in the vicinity of position 332 also contribute to target specificity (52, 54-56, 67). Several codon positions in *TRIM5* show elevated rates of non-synonymous substitution

and high levels of polymorphism (28), so arms races with retroviruses probably play out at multiple specificity-determining residues simultaneously. If multiple residues on a binding surface each engage in their own rock-paper-scissors chase, this will create a finite number of unique protein surfaces with unique interaction specificities. This has been demonstrated with primate protein kinase R (PKR), where amino acid combinations at three critical residues determine interaction with the poxvirus antagonist K3L (51). Importantly, this study and others demonstrate that the rules of engagement between two interacting proteins, even at a complex interaction surface, can be determined. Although the rock-paper-scissors model is almost certainly over-simplified, it may at least help begin to define constraints on viral evolution as discussed in more detail below.

Our knowledge of how viruses have responded over time to these cycling evolutionary forms of restriction factors remains weak. Alignments of viral genomes are usually limited to samples that have been collected within the last several decades, a scale that precludes analyzing mutational change through an evolutionary arms race. Importantly, the fact that viral genomes dating back tens of millions of years have now been found to be frozen in their historical form in the genomes of slowly-evolving animals may, for the first time, allow us the opportunity to look at viral adaptation over evolutionary scales. Nonetheless, some relevant data does currently exist. Rapid resampling of amino acids at certain residue positions has been observed in the HIV genome. This has been shown to reflect escape from MHC presentation in one host, and then due to the fitness cost of this escape mutation, reversion in another host (70). Importantly, this does support the idea that there are cognate viral forms that correspond

to different host immunity alleles. It was recently found that TRIM5α also restricts herpes simplex virus, but that this effect is highly virus strain-specific (71). This is again consistent with viral populations having cognate polymorphisms that exhibit escape from certain host alleles. Strain-specific interactions with host cells are commonly observed in the influenza field (72). Studies that utilize diverse clinical and laboratory isolates of a single virus are powerful and can reveal functional polymorphisms within viral populations, some of which may reflect individual adaptations to different host polymorphisms.

To reiterate, the rock-paper-scissors model describes a scenario where viral adaptation to a particular host genotype is highly constrained due to the delicate interplay of thousands of genetic determinants. Because of the limited number of genetic forms available, we have proposed a modified arms race model where both parties (host and virus) recycle a small number of alleles in a rock-paper-scissors chase. It remains unknown whether such dynamics will describe other systems. Thanks to high-throughput techniques, rapid progress is being made in understanding the host genetics of viral infection. New immunity genes and cofactors are still being discovered on a regular basis. The IFITM proteins, which restrict cellular entry of diverse viruses such as Dengue, West Nile, SARS, HIV, and Influenza A, were discovered only in 2009 (39, 73, 74). Signatures of positive selection have been documented in various host cofactor and immunity genes, suggesting that arms race dynamics will describe many other host-virus interactions (27, 75, 76). It remains to be seen how generally amino acid re-sampling, potentially reflective of rock-paper-scissors dynamics, occurs in these scenarios.

Predicting viral evolution, will it ever be possible?

It is tempting to speculate that the level of constraint observed may make forecasting of virus evolution possible. Important goals are to predict the evolution of viruses as they spread through populations, the transmission of viruses from one species to another, and the development of drug resistance. Due to the difficulty of these problems, the number of papers where prospective prediction of viral evolution has even been attempted remains small (77, 78). Many interesting evolutionary scenarios fundamentally require viruses to adapt to new host genotypic landscapes. The entire endeavor would be doomed if viral escape were truly unconstrained with no 'rules' to be found. However, the rock-paper-scissors model says that limited opportunities for escape are possible, at least if we assume that only point mutations will be utilized. Major and more rare genetic events utilized for escape, such as recombination or reassortment between viruses of different species, or the acquisition of novel viral genes, gene domains, or deletions within genes, will likely be impossible to predict. This may preclude prediction of evolution in viruses that experience these phenomena frequently, such as influenza.

Ultimately, forecasting viral evolution will require 1) a model of the selective constraints at play in any particular genotypic environment, and 2) knowledge of how the virus will respond to each of these constraints. Understanding the selective constraints imposed by different host genotypes will be difficult (see Outstanding Questions). This will require a comprehensive list of immunity and cofactor genes relevant to a particular

virus, a catalogue of the major alleles of each, and an understanding of which genes are the most potent barriers to, or facilitators of, infection. Next, different host alleles of each gene will each need to be characterized for their viral specificity. Viral escape from restrictive alleles will also need to be characterized, either empirically or using computational models, of the protein-protein interaction interfaces between host and viral proteins. Finally, all of this information has to be combined. Once such models exist, if a new genotypic environment presents to a virus one major genetic incompatibility in a known cofactor or immunity factor, and viral response to that block has been well characterized, prediction of viral evolution will be possible. Multiple genetic blocks to infection in a particular genetic background will require multiple corresponding adaptations of the virus, making prediction of viral evolution more complex but perhaps not impossible.

CONCLUSIONS

Evolutionary thinking in the HIV field has lead to many significant biological insights. For example, any given retrovirus is expected to be very well adapted to its natural host, and able to evade all aspects of that host's immune system. The opportunities for discovering novel mechanisms of resistance are limited in such systems. However, retrovirologists have clearly demonstrated that cross-species infection assays can lead to a rich description of host and virus genetics. By studying heterologous pairings of viruses and hosts, pairs that have not stayed 'in step' through arms race evolution, large genetic phenotypes of both immunity and virulence have been revealed. Such approaches are, at

their core, based on an appreciation for the antiquity of mammalian retroviruses and the fact that mammals have co-evolved with similar viruses for tens of millions of years. Now that we realize that many other virus families are ancient and ubiquitous in nature, such approaches should be applied in those fields. Reciprocally, such studies are likely to continue to refine the evolutionary theory of host-virus interactions.

OUTSTANDING QUESTIONS

- What is the global repertoire of host antiviral and proviral genes that interact with each virus?
- Which host loci present the most potent genetic barriers to cross-species and individual-to-individual virus transmission?
- How many functionally different alleles circulate at each proviral and antiviral locus in host species, and how are they geographically localized?
- Once restrictive cofactor and immunity alleles are known, how do different viral isolates escape them?
- Once relevant host genes have been described, and viral escape to each defined, can models be derived that predict viral evolution in a new host?

Chapter 2*

The effect of species representation on the detection of positive selction in primate gene datasets

HIV arose from a family of viruses that has infected non-human primates for millions of years. During this time, both host- and virus-encoded genes have been continually selected to modify their interactions with one another. This has resulted in the rapid evolution of the specific codons that govern the physical interactions between host and virus proteins. Virologists have discovered that these evolutionary signatures, acquired in nature, can provide a shortcut in the functional dissection of these host-virus interactions in the laboratory. However, the use of evolution studies in this way is complicated by the fact that many non-human primate species are endangered, and biomaterials are often difficult to acquire. Here, we assess how the species representation in primate gene datasets affects the detection of positive natural selection. Our results demonstrate how targeted primate sequencing projects could greatly enhance research in immunology, virology, and beyond.

* Ross McBee, Shea Rozmiarek, Dr. Paul Rowley, and Dr. Sara L. Sawyer all contributed to the conception of this project, data collection, and data analysis. This work has been accepted for publication and is in press at Molecular Biology and Evolution. Permission to adapt the contents of the publication was acquired from the co-authors.

INTRODUCTION

The evolution of human genes can be studied in multiple ways (79). One approach uses comparisons of gene orthologs from humans and nonhuman primates to analyze selection over long evolutionary timescales. The main metric used is dN/dS, which summarizes the rates of non-synonymous (dN) and synonymous (dS) DNA substitutions in gene sequence (80-82). Codons where dN/dS > 1 have experienced natural selection in favor of non-synonymous mutations. Viruses and their hosts co-evolve over long periods of time and, as a result, the dN/dS > 1 signature can often be detected in gene regions corresponding to physical interaction interfaces between host and virus proteins. The identification of codons with dN/dS > 1 in primate genes has recently become important in guiding genetic studies in the HIV field, having been particularly powerful in dissecting the physical interactions of several human immunity proteins with HIV (17-20, 83-86). The identification of this evolutionary signature has been recognized as a short-cut in the laborious functional dissection of host-virus interactions, and is now being used to characterize the interplay between host proteins and other types of pathogens as well (25, 51, 75, 87-91).

The main limitation to using this analysis is the acquisition of appropriate primate sequence datasets. There are currently nine available simian primate genome projects available through the UCSC genome browser (http://www.genome.ucsc.edu/), but the acquisition of additional primate sequences is complicated by the fact that many nonhuman primate species are endangered, and the purchase of immortalized cell lines derived from these species can require a federally-issued permit. Because the analysis of

dN/dS has become so useful in guiding the genetic study of human genes that interact with viruses, we wished to investigate how many primate sequences are required to reliably detect positive selection.

MATERIALS AND METHODS

Primate sequence datasets

All sequence alignments were previously generated by our lab (27, 92).

PAML analysis

Each gene alignment was fit to two codon models, M7 and M8, as implemented in PAML (93). A likelihood ratio test was then performed, using 2 degrees of freedom, to assess whether M8 (permitting some codons to evolve under positive selection) gives a significantly better fit to the data than M7 (positive selection not allowed). The Bayes Empirical Bayes approach was then used to calculate posterior probability that each codon is properly assigned to the dN/dS > 1 site class (94). This entire protocol was performed for each of the 11 datasets, for each of the subtrees analyzed.

Generation of random trees

The analysis shown in panels D-F of Figure 2-2 required the generation of random trees. Ten random trees of each of the 7 tree sizes ("number of species" on X axis) were generated for a total of 70 unique trees for each gene. To generate the random trees, each species in each of three primate clades (color coded in Figure 2-1A) was

assigned an integer and random.org was used to generate 10 unique sets of non-repeating integers. For the first 4 of these sets of trees, 1, 2, 3, and then 4 species were chosen at random from each of the 3 primate clades resulting in trees containing a total of 3, 6, 9, and 12 species. However, because sequences were available from only 4 New World monkeys, tree sets past the 12-species size only introduced two new species at each step, one from each of the two other clades.

RESULTS

To test the sensitivity of positive selection analyses when using different primate datasets, we re-analyzed datasets that our group has previously generated for 11 different genes (*XLF, XRCC4, MAP4, NBS1, CtIP, WNK1, POLλ, NUP153, RANBP2, IBTK*, and *NUP98/96*) (27, 92). These 11 datasets each consist of 20 orthologous sequences from a matched set of primate species (Figure 2-1A). In the studies where these datasets were generated, genes were chosen for sequencing based on higher-than-normal levels of protein divergence, or because they are known to encode proteins important for viral lifecycles. Therefore, this is not a random set of genes, but rather a set that is skewed towards positive selection.

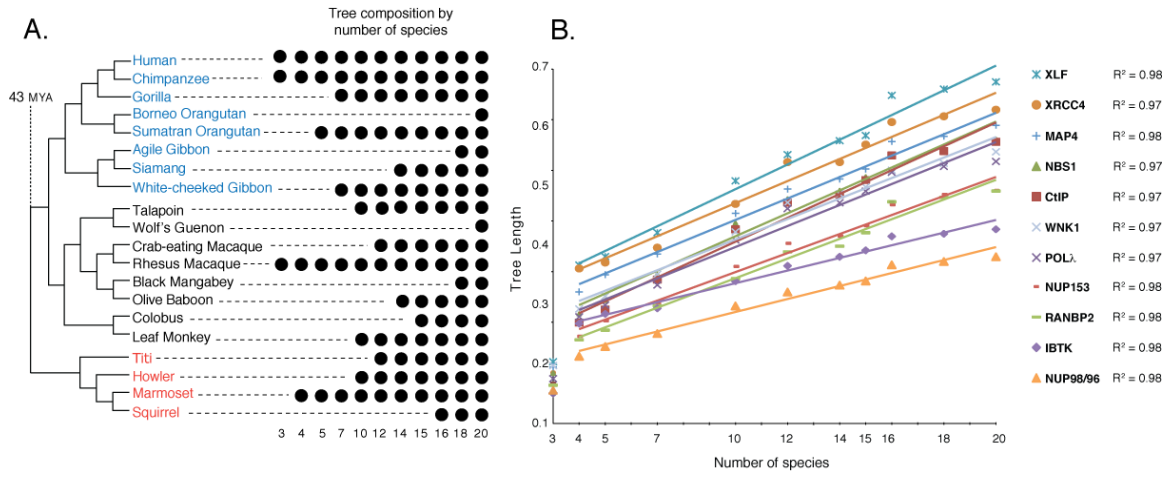


Figure 2-1: Primate datasets representing different levels of divergence. (A) The master tree of 20 species from which subsequent "pruned" trees were derived, as well as a matrix showing which primates were included in those pruned trees. Simian primates are broken into three major groups: hominoids/apes (blue), Old World monkeys (black) and New World monkeys (red). **(B)** The overall divergence in each dataset (shown in tree length) as the number of species increases. Tree length is the sum of all branch lengths on a tree. A line is fit to the data for each gene (R^2 values reported in the legend), excluding the three species tree because the tree length is so low for this set.

We then generated 10 pruned trees representing subsets of these 20 species (Figure 2-1A). The first four trees were made to reflect the history of primate genome sequencing projects. For instance, the first tree that we analyzed was a 3-species tree of human, chimpanzee, and rhesus macaque, representing the first three primate genomes sequenced (43, 49). The 4-species tree also included marmoset, the fourth primate to have its genome sequenced. We then added the fifth sequenced primate species, Sumatran orangutan (95), creating a 5-species tree, and then white-cheeked gibbon and gorilla (96), creating a 7-species tree. We specifically chose this strategy to evaluate the power of early positive selection studies that only had access to a limited number of primate genomes (49, 95, 97-99). Beyond this, 10-, 12-, 14-, 15-, 16-, and 18- species subtrees

were made (Figure 2-1A). The species included in these trees were chosen so that "tree length" scales approximately linearly with the number of species included (Figure 2-1B). Tree length is the sum of the branch lengths along the tree or, in other words, the average number of nucleotide substitutions per site in an alignment. For any given tree, we find higher tree length in some of our datasets than others (Figure 2-1B). For instance, the DNA repair gene *XLF* has the highest level of sequence divergence, and the nuclear pore gene *NUP98/96* has the lowest level of divergence. This set of trees was then used to test the effects of species representation on the detection of positive selection.

Because the PAML software suite is now commonly used in virology research, we focused on the detection of positive selection using codon models implemented in PAML's codeml program (93, 100). To detect selection, each alignment was fit to the codon models M7 (null model, codon values of dN/dS fit to a beta distribution bounded between $0 < dN/dS < 1$) and M8 (positive selection model, similar to M7 but with one extra site class assigned at $dN/dS > 1$). A likelihood ratio test was then used to determine whether the null model (M7) could be rejected in favor of the model of positive selection (M8). We performed likelihood ratio tests between M7 and M8 for all 11 gene datasets, using each of the 11 possible trees. These genes fell into three classes. First, six genes converged on significant rejection of the null model ($p < 0.05$) as more species were added (Figure 2-2A). One of these, *XRCC4*, reached significance after only 4 species, and stayed significant as more species were added. On the other extreme, *MAP4* did not reach significance until the 20 species dataset. We conclude that more species allow an increased possibility of detecting positive selection, although some genes are more

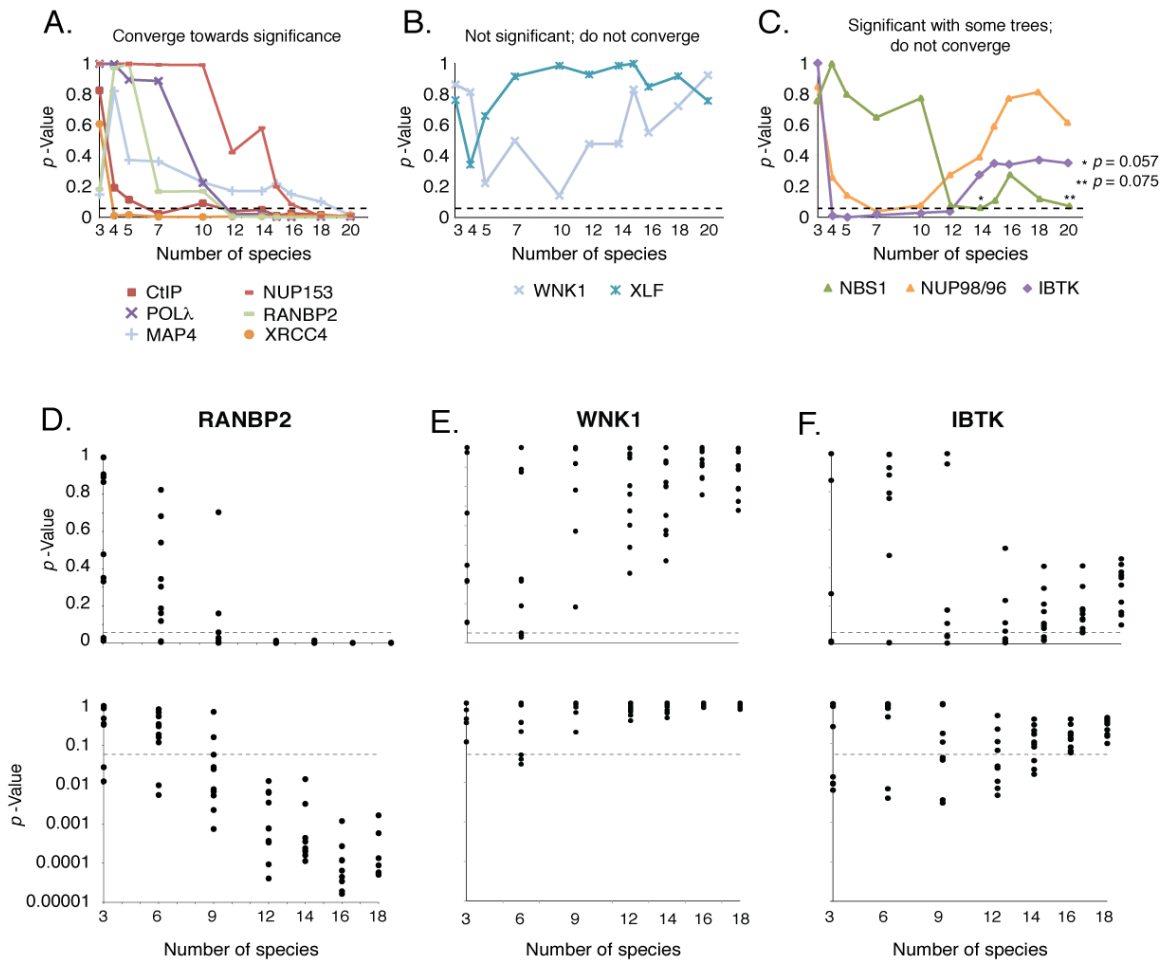


Figure 2-2: The impact of dataset composition on PAML's ability to detect positive selection. (A-C) Each point represents a single model comparison performed. The x-axis denotes the primate tree that was used, as defined in Figure 2-1A, and the y-axis is the calculated *p*-value of the M7-M8 likelihood ratio test. The dashed line indicates a significant *p*-value ($p < 0.05$) where the null model is rejected. Panel A shows those genes that converge on a significant *p*-value as more primate species are added. Panel B shows those genes that do not. Panel C shows genes that lack a clear convergence towards a stable *p*-value as more species are added. **(D-F)** These graphs each represent data for a single gene from panels A-C. In this case, 10 alternate species sets were randomly chosen for each number of species shown on the X-axis. Each of these random trees was used to evaluate M7 and M8, and the *p*-value is calculated for each model comparison. The Y-axis is shown on a linear (top) and log (bottom) scale.

sensitive than others. A second set of 2 genes (*WNK1* and *XLF*) did not reach the $p < 0.05$ significance threshold using any of the trees tested (Figure 2-2B). Further, for these genes there is no clear trend towards significance, suggesting that the null hypothesis would never be rejected even if more sequences were added. Finally, for a third set of 3 genes, the null hypothesis is rejected (or very nearly so in the case of *NBS1*) with smaller datasets, but then the likelihood ratio test loses significance as more species are added (Figure 2-2C). These genes might be experiencing positive selection specifically in the hominoid clade. The smaller datasets are hominoid-rich because several of the first species sequenced were primates closely related to humans (e.g. chimpanzee and orangutan). In fact, we have previously substantiated hominoid-specific positive selection for one of these three genes, *NBS1* (92). Large scale screens for positive selection using fixed species sets will likely miss these patterns, and special tests will need to be run to test alternate hypotheses like clade-specific positive selection. In summary, except for genes with clade-specific effects, the 20 species dataset that we show here is adequate for detecting selection in all of the genes tested. For genes that do not reach significance by the time 20 species are included, there is no trend indicating that the addition of more sequences will lead to the rejection of the null hypothesis.

In the analyses just discussed, there is stochastic noise in the patterns observed. Based on this, we wished to test how likely it would be to get a false signature of positive selection. We next examined more closely the effects of primate species choice on the evolution of one gene from each of the three classes of evolution that were just described (*RANBP2*, *WNK1*, and *IBTK*). Likelihood ratio test results for 3-, 6-, 9-, 12-, 14-, 16-, and

18- species trees were again examined, but this time randomly chosen species from the 20-species collection were used to create 10 trees of each of these sizes. Those trees and the corresponding sequences were then fit to M7 and M8, and the *p*-value of the likelihood ratio test is shown (Figure 2-2D-F). With small, three species trees, a broad range of *p*-values, ranging from $0 < p < 1$, were observed. In other words, the results were highly stochastic and depended on the specific three species chosen for analysis. As more species were added to the tree, the variance of these results diminished and converged on a true value. This trend continues well below the statistical cutoff of $p = 0.05$, as seen when using a log scale (Figure 2-2D-F bottom). For instance, for RANBP2, all 10 randomly generated 12-species trees have $p < 0.05$, and this is true for all trees larger than 12 species as well (Figure 2-2D). Thus, it is possible to get a false signature of positive selection due to stochastic effects, but that the likelihood of this decreases as more species are included.

If the null model (M7) is rejected in favor of the model of positive selection (M8), the individual codons assigned to the dN/dS > 1 bin can be used to guide genetic studies (17-20, 83-85, 101). The logic is that non-synonymous mutations in these codons impact function, otherwise selection would not be acting on these sites. For the 6 genes that pass the likelihood ratio test (Figure 2-2A) we examined how increasingly rich datasets affect the fraction of codons assigned to the dN/dS > 1 class in M8. In general, this value stabilizes by the these 20 simian primate species were included (Figure 2-3A). The proper assignment of each codon to this class can also be evaluated using a posterior probability (94). If a codon has a posterior probability of $P \geq 0.95$, there is a 95% chance that this

codon is correctly assigned to the dN/dS > 1 class. We then looked at the fraction of sites in each dataset assigned to the dN/dS > 1 class with a posterior probability ≥ 0.95 (Figure 2-3B). These values also stabilize after inclusion of these 20 simian primate species. The specific codons identified are illustrated for two of these genes, RANBP2 (Figure 2-3C) and Polλ (Figure 2-3D). Interestingly, it appears that many sites identified with as few as 4-5 sequences, even when the likelihood ratio test has low statistical support (Figure 2-2A), are often increasingly supported as more sequences are added. This surprising finding suggests that it may be worth functionally testing codons identified even if only a few sequences are available for analysis, and even before the likelihood ratio test has reached the rigorous $p < 0.05$ value.

DISCUSSION

In conclusion, positive selection can be adequately characterized in primate genes with a 20-species dataset composed of 8 hominoids, 8 Old World monkeys, and 4 New World monkeys (Figure 2-1A). Tree length, a measure of overall divergence in an alignment, can also be used as a guide, and 20-species datasets have tree lengths between approximately 0.3 and 0.65 (Figure 2-1B). This study should serve both evolutionary biologists and virologists who are interested in the molecular evolution of genes during the course of simian primate speciation. It also underscores the need for more sequenced primates genomes, which would alleviate the burden on individual researchers to obtain these precious primate biomaterials.

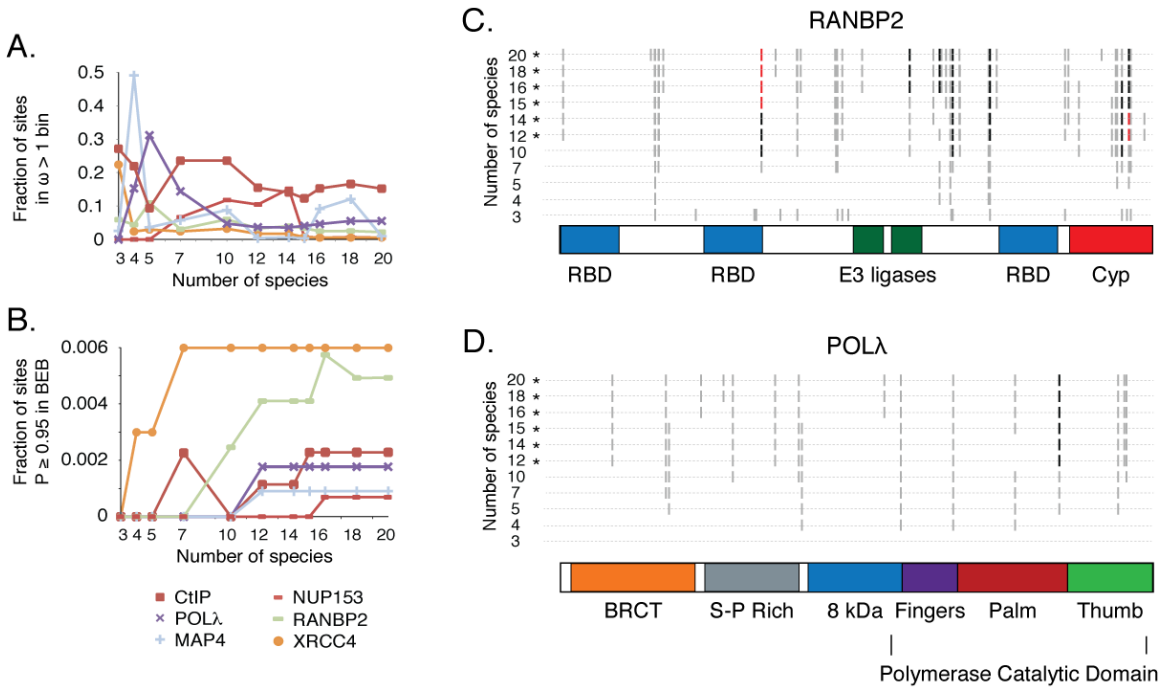


Figure 2-3: The impact of dataset composition on the identification of codons targeted by positive selection. **(A)** The fraction of the codons in each gene that were placed in the $\omega > 1$ bin in M8, using datasets composed of increasing numbers of species. **(B)** The fraction of codons in the $\omega > 1$ bin in M8 with a posterior probability of $P \geq 0.95$. **(C and D)** Domain diagrams for RANBP2 and Polλ with the locations of codons, shown as dashes, placed in the $\omega > 1$ bin in the M8 model with a posterior probability ≥ 0.5. Codons with a posterior probability of $P \geq 0.95$ are highlighted in black, and those with posterior probability of $P \geq 0.99$ in red. Each row represents an analysis performed with a different number of species, with asterisks indicating datasets for which the null model (M7) is rejected by the likelihood ratio test ($p < 0.05$).

Chapter 3*

HIV-1 host factors identified through evolutionary analysis

HIV-1 requires the functions of many human proteins to replicate in human cells. Because these proteins constitute viable targets for antiviral drugs, there has been intense interest in identifying and characterizing these proteins. Genome-wide screens have generated large numbers of new, putative HIV-1 host factors, and the next challenge is to prioritize these candidates for in-depth study. Towards this goal, we made use of an evolutionary signature, caused by recurrent positive selection, which is commonly found in host genes encoding virus-interacting proteins. Here, five genes from these screens that bear the signatures of positive selection were identified. Some of these genes (*CD4, NUP153, RANBP2/NUP358)* are well characterized with respect to the HIV-1 lifecycle, whereas others (*ANKRD30A/NY-BR-1* and *MAP4)* remain relatively uncharacterized. We find that ANKRD30A, which has never been functionally studied in the context of the HIV-1 life cycle, interacts through its C-terminal domain with HIV-1 soon after the virus enters cells. These data provide a rationale for why ANKRD30A was identified in only a subset of high-throughput screens for HIV-1 host factors. This addresses an outstanding issue in the interpretation of the large, genome-wide datasets being generated in virology

* Drs. Paul A. Rowley and Christina H. Swan assisted with sequencing primate genes. Dona T. Le prepped nucleic acid from primate blood samples. Dr. Gregory K. Wilkerson supplied primate blood samples. This work was published in Virology [27]. Permission to adapt the contents of the publication was acquired from the co-authors and the publishing company.

research: the often low overlap between them. We propose that evolutionary analysis can be a powerful addition to the systems biology-based models of human-virus interactions that are now required.

INTRODUCTION

HIV-1 exploits a vast network of human proteins to replicate within human cells. These human proteins, referred to here as HIV-1 host factors, are involved in processes such as transcription, translation, and transport of the virus (102-105). Experimentally, the depletion or disabling of HIV-1 host factors decreases viral replication. For this reason, these host proteins constitute novel and potentially highly effective antiviral drug targets. The FDA has recently approved the first anti-HIV drug targeting such a protein, Maraviroc, which is a small molecule antagonist of the HIV-1 co-receptor CCR5. For this reason, there is intense interest in human proteins that promote the replication of viruses that cause human disease, including HIV-1, influenza, West Nile virus, Dengue virus, and others (reviewed in (23, 76, 106-111)).

A current assessment of the literature would suggest that HIV-1 requires hundreds, or potentially more than a thousand, human proteins to facilitate its replication inside of the cell. In 2008, a survey of HIV-1 host interactions reported in the HIV-1 literature led to the creation of the "HIV-1, Human Protein Interaction Database" hosted by NCBI (112, 113). In that survey, 1,448 human proteins were identified that are either bound, inhibited, upregulated, or modified by HIV-1. As there are only 15 distinct HIV-1 proteins, this equates to 96 direct or indirect interactions, on average, for every viral

protein. Since that time, high-throughput screens for HIV-1 host factors have been conducted. In several genetic screens, RNA interference was used to systematically deplete human gene transcripts, and then the effects of depletion on HIV-1 replication were evaluated (reviewed in (114)(22)). Collectively, these screens uncovered approximately a thousand genes that lead to reduced HIV-1 replication when their expression is decreased (115-118). Another proteomics-based study recently employed affinity purification and mass spectrometry to uncover 435 human proteins that interact with HIV-1 proteins (119). Although these high-throughput technologies have undoubtedly led to new insights, one surprise has been the low overlap between the different screens, and between the screens and the interactions previously reported in the literature (22, 120). Similar scenarios have played out in other fields of virology where high-throughput studies are being employed (106)(121). Despite many important new host factor genes being identified, the task now at hand is to triage these gene lists for further, more in depth mechanistic studies.

To achieve this goal, we took advantage of a unique evolutionary signature that has been faithfully found in all known HIV-1 restriction factors (17-20, 38, 84, 85, 122, 123). This evolutionary signature results from the millions-of-years long struggle for survival between retroviruses and the primates that they infect (124-129). In contrast to host factors, restriction factors block the replication of HIV-1 upon recognition of and interaction with specific viral targets (7, 11, 104, 130). Retroviruses, in turn, encode antagonist proteins that specifically recognize and inhibit some restriction factor proteins

(131). The evolutionary battles between restriction factors and viruses play out at physical interaction interfaces between host and virus proteins. Both parties (host and virus) are continuously selected for mutations that modulate this interaction to give each the upper hand over the other. For instance, the TRIM5α restriction factor has experienced continuous selection to better recognize its target, the retroviral capsid (17), whereas capsid continuously evolves to escape interaction with TRIM5α (28)(132). This continual evolutionary struggle is referred to as an evolutionary "arms race" and, because it occurs at the level of protein-protein interactions, results in the rapid evolution of each of the interacting host and viral proteins (127, 133). This signature of rapid evolution is so typical of restriction factors that it has even be used to predict novel restriction factors (38, 133-136), as well as the virus-binding domains of restriction factors (17, 126, 127, 133).

Although the application of these evolutionary analyses to the study of HIV-1 restriction factors is now common practice, we hypothesized that HIV-1 host factors evolve under positive selection due to their long-term evolutionary engagement with viral proteins. In this scenario, virus genomes would be selected for mutations that improve physical interaction between virus proteins and beneficial host factors. In turn, host genomes would be selected for mutations in these host factors that reduce interactions with viruses. However, it is not known whether HIV-1 host factors will have the evolutionary flexibility to engage in arms race dynamics the way that restriction factors do. Unlike restriction factors, which are typically dedicated proteins of the immune

system, host factors that facilitate viral replication tend to play important roles in cellular physiology and so would be expected to have significantly more evolutionary constraints acting upon them. Thus, while these proteins might easily be able to accrue mutations that would allow them to avoid being hijacked by viral pathogens, the question is whether they will be able to do so without major collateral damage to the cell through loss or alteration of their housekeeping functions. Several recent reports indicate that virus host factors also experience positive selection and arms race dynamics (25, 89, 90, 92, 137-139). This idea was tested further by harnessing the predictive power of this evolutionary signature to help highlight interesting new HIV-1 host factors uncovered in high-throughput screens.

Patterns of molecular evolution were analyzed in all of the human genes that have been identified in two or more independent RNA interference screens for HIV-1 host factors (115-118). This analysis required the generation of 160 primate gene sequences. We find that HIV-1 host factor genes are, overall, more conserved than restriction factor genes, as would be predicted by the higher level of evolutionary constraint acting on them. Nonetheless, these genes can experience positive selection at certain residue positions, just as restriction factors do. Positive selection was detected in three human genes with well-established roles in promoting the HIV-1 lifecycle (*CD4, NUP153, RANBP2*) and two less characterized genes (*ANKRD30A* and *MAP4)*. The identification of these well-characterized host factors supports the idea that positive selection analysis can be a powerful addition to large-scale screening efforts, and indicates that the

remaining two genes, *MAP4* and *ANKRD30A*, are worthy of closer examination. We verify ANKRD30A as an HIV-1 host factor, showing that this protein interacts with HIV-1 through its C-terminal domain at a stage between virus entry and nuclear import. Finally, a potential rationale for why ANKRD30A has been identified in some high-throughput screens for HIV-1 host factors but not others is presented, directly addressing why candidate proteins found in high-throughput genetic and proteomic screens often show limited overlap.

MATERIALS AND METHODS

Primate biomaterials

Primary and immortalized primate cell lines from primate species were grown in standard media supplemented with 15% fetal bovine serum at 37ºC and in 5% CO_2. Macaque, owl monkey, and squirrel monkey samples included in the population study were acquired from either the Michale E. Keeling Center for Comparative Medicine and Research (Bastrop, Texas), or from the New England Primate Research Center (Southborough, MA). For these individuals, 2.5 mL of whole blood was collected in PaxGene Blood RNA Tubes (BD, #762165). Alternately, B-cell lines were expanded in suspension culture, in RPMI, 20% FBS, Pen/Strep, L-glutamine, HEPES, and AZT. Genomic DNA and RNA from primate blood and cell lines was isolated using the PaxGene miRNA kit (Qiagen, #763134) and/or the Qiagen All Prep DNA/RNA mini kit (Qiagen, #80204).

Primate gene sequences

Human Refseq sequences were obtained from the NCBI nucleotide database. Chimpanzee, orangutan, rhesus macaque, and marmoset gene sequences were obtained from the UCSC genome database (http://genome.ucsc.edu/) using the BLAT alignment tool. Genes were sequenced from additional primate species, and from chimpanzee, orangutan, rhesus and marmoset in instances where the genome-project sequences were of poor quality. PCR or RT-PCR was performed from total RNA, gDNA, or cDNA with SuperScript III One-Step RT-PCR system with Platinum Taq (Invitrogen, #12574-018), PCR SuperMix High Fidelity (Invitrogen, #10790-020), or Phusion High Fidelity PCR Master Mix (NEB, #F-531S). Primate gene sequences have been deposited in GenBank (accession numbers KJ531711-KJ531825).

Human polymorphism analysis

Human SNPs were identified in datasets deposited by the 1000 Genomes Project (http://browser.1000genomes.org, release 13) and NCBI's dbSNP database (http://www.ncbi.nlm.nih.gov/SNP).

Sliding window analysis

Sliding-window dN/dS calculations for each alignment were performed with the SLIDERKK program (140). Human-orangutan, human-rhesus and rhesus-marmoset alignments were analyzed with standard window sizes (92) of 450bp, 306bp and 153bp, respectively, to reflect the increasing level of divergence in these species pairs (window

size must be a multiple of nine in this program). To generate confidence values for windows with dN/dS > 1, the K-estimator program (141) was utilized to generate a null distribution of dN/dS values through Monte Carlo simulation in the gene region of interest.

PAML analysis

Codon models were tested with codeml in the PAML 4.1 software package (93). To detect selection, multiple alignments were fit to the NSsites models M8a (neutral model, codon values of dN/dS fit to a beta distribution plus an extra codon class fixed at dN/dS = 1) and M8 (positive selection model, similar to M8a but with the extra class allowed to be dN/dS >1). A likelihood ratio test was performed to assess whether permitting codons to evolve under positive selection gives a significantly better fit to the data (model comparison M8a vs. M8). In situations where the null model could be rejected ($p < 0.05$), posterior probabilities were assigned to individual codons belonging to the class of codons with dN/dS > 1. For the whole-gene dN/dS values indicated on the genome-wide distribution of human/chimpanzee/rhesus gene trios, dN/dS values were calculated using the M0 model in PAML.

ANKRD30A and MAP4 cloning

The *ANKRD30A* gene fragment was synthesized and cloned into the plasmid pUC57 (synthesis and cloning by GenScript). The *MAP4* gene fragment was amplified from human cDNA (Clontech, #636643) and TA-cloned into pCR4 (Invitrogen, #K4575-

01). For both of these genes, an N-terminal 3xFLAG tag was added using a PCR reaction and these tagged constructs were TA-cloned into the gateway entry plasmid pCR8 (Invitrogen, #K2500-20). An LR Clonase II reaction (Invitrogen, #11791-100) was used to move these constructs into a Gateway-converted pLPCX retroviral vector (Clontech, # 631511).

Generation of stable cell lines

To produce cell lines that stably express MAP4 and ANKRD30A fragments, retroviral vectors were used to transduce CRFK cells (ATCC). 293T cells were seeded at a concentration of $1x10^6$ cells/well in a 6-well dish. After 24 hours each well was transfected with 2 μg of pLPCX construct (empty or encoding the gene fragment of interest), 1 μg pCS2-mGP encoding MLV gag-pol (142), and 0.2 μg pC-VSV-G at a final 1:3 ratio of DNA to Fugene (μg DNA : ml Fugene6). Supernatants were collected after 48 hours, passed through a 0.2 μm filter, and used to infect CRFK cells grown in DMEM supplemented with 10% fetal bovine serum, 100 IU/ml penicillin, 100 μg/ml streptomycin, and 2 mM L-glutamine. After 24 hours, media containing 8 mg/ml puromycin was added to select for transduced cells. Expression of ANKRD30A and MAP4 constructs was detected by Western blot.

Western blot analysis

Puromycin-resistant cell lines were grown to confluency in a 6-well dish, collected using a cell scraper, and lysed in a buffer containing 150 mM NaCl, 50 mM

Tris-HCl (pH 7.4), 1% NP-40, and Complete protease inhibitor (Roche, #11836170001). After quantitation of protein concentration using a Bradford assay, 30 μg (ANKRD30A) or 1 μg (MAP4) of whole cell extract was resolved using a 10% polyacrylamide gel and transferred to a nitrocellulose membrane. FLAG-tagged constructs were detected using a 1:2000 dilution of mouse anti-FLAG antibody 3B9 (Syd Labs, #PA000274-M20008L). A 1:10,000 dilution of donkey anti-mouse horseradish peroxidase-conjugated antibody (Thermo Scientific, #32430) was used as a secondary probe. β-actin was also detected as a loading control using a 1:1000 dilution of mouse anti β-actin (Santa Cruz, #sc-47778). Blots were developed using the ECL Plus detection reagent (GE Healthcare, #RPN2132).

Microscopy preparation and confocal imaging

CRFK cells stably expressing FLAG-tagged fragments of MAP4 or ANKRD30A, or transduced with an empty vector, were seeded into a 4-well permanox chamber slide (LabTek, #70400) at a concentration of 100,000 cells/well. Cells were fixed with 4% paraformaldehyde in PBS for 10 minutes, permeabilized with 0.5% Triton-X in PBS for 15 minutes, blocked with 3% bovine serum albumin (Sigma-Aldrich, #A7906-50G) in PBS for an hour, stained with a 1:500 dilution of anti-FLAG antibody (Syd Labs, #PA000274-M20008L), and then visualized using a 1:1000 dilution of Alexa Fluor 594 antibody (Molecular Probes, #A-11005). Cells were mounted using Vectashield mounting media containing DAPI stain (Vector Laboratories, #H-1200). All incubations were done at room temperature and followed by three washes using 3% bovine serum albumin in PBS. Confocal microscopy was performed using a Zeiss LSM 710 confocal

microscope with a 40x objective. All images were collected with identical laser and exposure times. Confocal slices of 1 mm were taken in the nuclear volume. Images were analyzed using the ImageJ version 1.43u software package (143).

Retroviral integration assays

Viruses for single-cycle infection assays were packaged in 293T cells by co-transfection of plasmids encoding viral proteins and VSV-G, along with a transfer vector, as follows: HIV-1 (pMDLg/pRRE, pRSV-Rev, pMD2.G, pRRLSIN.cPPT.PGK-GFP.WPRE; all available from Addgene), FIV (pFP93 (144), pC-VSV-G, pGIN-SIN:GFP (144)), NB-MLV (pCS2-mGP, pC-VSV-G, pLXCG (142)). After 48 hours, supernatant containing viruses was harvested, filtered, and frozen. For infection assays, CRFK stable cells lines were plated at a concentration of $5x10^4$ cells/well in a 24-well plate and infected with HIV-1, FIV, or NB-MLV the following day. Two days post-infection, cells were analyzed by flow cytometry for expression of GFP using the BD Bioscience Fortessa cell analyzer. Dose curves were performed in triplicate, and results were confirmed with independent experiments.

Expression analysis of *ANKRD30A* in human tissues and laboratory cell lines

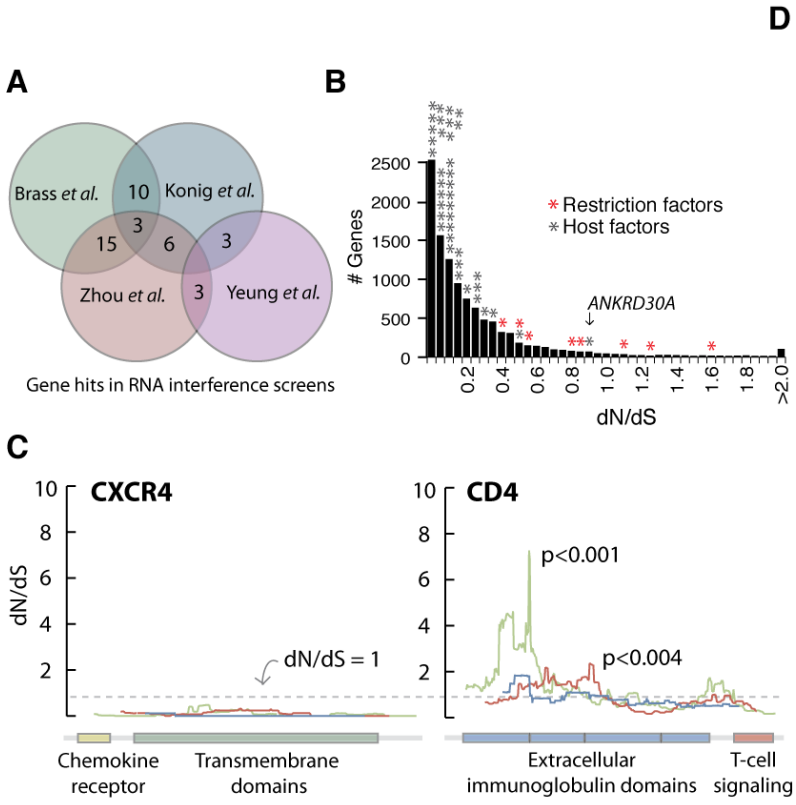
Primers were designed to amplify a 750bp fragment that spans the last two exons of *ANKRD30A*. PCR reactions were performed with PCR SuperMix High Fidelity (Invitrogen, #10790-020) along with primers NRM602 (forward: 5'-TTAGGGAAGAATTAGGAAGAATC-3') and NRM604 (reverse: 5'-

CATTTGACACTGTGTTTCACGTTG-3'). The Human Total RNA Master Panel II (Clontech, #636643) was converted to cDNA using the SuperScript III First-Strand Synthesis System (Invitrogen, #18080-051) and used as a template in PCR reactions. The Human MTC cDNA Panel II (Clontech, #636743) was also utilized. HeLa, 293T, Jurkat, and breast carcinoma (HCC1937) cells were grown in DMEM and RNA was harvested using an All Prep DNA/RNA mini kit (Qiagen, #80204). Each of these cell lines was also treated with IFN- β (Betaseron) at a concentration of 1000 units/ml for 24 hours prior to isolation of RNA. *ANKRD30A* transcripts detected were verified via sequencing.

RESULTS

Five putative HIV-1 host factors have evolved under positive selection

In this study, we concentrated on the 40 human genes that have been identified in two or more of the large-scale RNA interference screens for HIV-1 host factors (Figure 3-1A) (22, 117). Positive selection can be detected by analyzing the "dN/dS" ratio, which summarizes the rate at which non-synonymous (amino-acid altering; dN) and synonymous (not altering the encoded amino acid; dS) mutations have accumulated in a gene over evolutionary time. Repeated rounds of positive natural selection for non-synonymous mutations result in a characteristic signature of dN/dS > 1 because non-synonymous mutations tend to be more deleterious than synonymous mutations. First, dN/dS in HIV-1 host factor and restriction factor genes were compared. Figure 3-1B shows a distribution of whole-gene dN/dS values for over 10,000 primate genes (49).



D

Gene	Human-Orangutan	Human-Rhesus	Rhesus-Marmoset
ADRBK1	.51	.10	.17
AKT1	.18	.13	.15
ANAPC2	.16	.39	1.42
ANKRD30A	2.05*	5.60*	6.62*
CAV2	.87	.81	1.22
CCNT1	.89	.33	1.46
CD4	1.88	2.53*	7.38*
CHST1	.00	.00	.12
CTDP1	1.07	.60	.81
CXCR4	.12	.35	.46
DDX3X	.10	.08	.13
DMXL1	1.70	2.37*	1.88
EPAS1	.97	.25	.00
HMCN2	1.53	1.99	3.17*
IBTK	1.71	2.20*	3.80*
IDH1	.48	.60	.82
JAK1	.49	.28	.85
MAP4	2.02*	1.50	6.18*
MED4	.00	.11	.19
MED6	.55	.27	1.02
MED7	.00	.00	.35
MED14	.46	.46	1.52
MED19	.12	.08	.00
MED28	.00	.09	.54
MID1IP1	.00	.00	.32
MRE11A	.65	.73	2.05*
NFKB1	.49	1.06	2.85*
NUP98/96	2.16*	2.05*	1.49
NUP153	4.22*	1.05	2.39*
RAB11A	.00	.16	.00
RAB28	.51	.24	.46
RANBP2	4.47*	3.17*	2.49*
RELA	.46	.45	1.22
RNF26	.36	1.56	.73
RPL3	.00	.00	.10
TCEB3	.36	2.53*	.93
TNPO3	.00	.40	.48
TRIM55	2.57*	.39	.99
WNK1	2.17*	3.52*	2.05*

Figure 3-1: Evolutionary analysis of genes identified in at least two RNA interference screens for HIV-1 host factors.

Figure 3-1: Evolutionary analysis of genes identified in at least two RNA interference screens for HIV-1 host factors. (**A**) Each circle represents a whole-genome RNA interference screen previously conducted to identify human genes important for HIV-1 replication (115-118). Within each overlap region is listed the number of human genes identified in multiple screens. Only three human genes (*RELA, MED7, MED6*) were identified in three screens, and no genes were identified in all four screens. (**B**) A histogram shows whole-gene dN/dS values previously calculated for 10,376 orthologous gene trios from the human, chimpanzee, and rhesus genomes (49). On the x-axis is shown the average dN/dS calculated over the length of each gene for this three-species tree. On the y-axis is the number of human-chimp-rhesus gene trios with this average dN/dS value. On this distribution are overlayed the dN/dS values for similar trios hand-curated for known HIV-1 restriction factors (red asterisks), and for the 40 putative host factors analyzed in this study (gray asterisks). The restriction factors are (from left to right on the distribution): *SAMHD1, Tetherin, ZAP, TRIM22, TRIM5, APOBEC3H, APOBEC3DE, APOBEC3G*. (**C**) Example sliding window analyses for *CD4* and *CXCR4*. The x-axis is in base pairs and represents the length of each gene. dN/dS was calculated in sliding windows moving along the length of the gene. The three pairwise species comparisons made are color-coded (blue, human vs orangutan; red, human vs rhesus; green, rhesus vs marmoset). For each comparison, the gene region with the maximum dN/dS value was tested for statistical significance, and regions where dN/dS is significantly greater than 1 ($p < 0.001$ and $p < 0.004$) are indicated. (**D**) The maximum dN/dS value observed in each pairwise sliding window analysis is summarized. An asterisk (*) indicates instances where $dN/dS > 1$ ($p < 0.05$).

As expected, most primate genes have dN/dS << 1. Overlayed on this distribution are the values of the 40 host factor genes analyzed in this study (gray stars). For comparison, dN/dS values for known restriction factor genes are also shown (red stars). These data show that genes encoding host factors are, in general, more conserved than genes encoding restriction factors, consistent with previous reports (99) and with their role as housekeeping genes.

In host genes subject to arms races with viruses, patterns of dN/dS > 1 would not be expected to be evenly distributed throughout the entire length of a gene, but rather to be specifically concentrated in gene regions correlating to the protein-protein interaction surface with the viral antagonist (127). Rather than analyzing dN/dS on a whole-gene basis, gene regions that are experiencing positive selection were determined. For each of the 40 genes being analyzed, sequences from three species with available genome projects were used to construct three pairwise gene alignments (human vs orangutan, human vs rhesus, and rhesus vs marmoset). For each of these alignments, dN/dS was calculated in sliding windows along the length of each gene (38, 92, 138). Examples of the sliding window results are shown for two of the 40 genes analyzed, *CD4* and *CXCR4*, encoding two receptors for HIV-1 entry (Figure 3-1C). Typical of most genes in primate genomes, dN/dS << 1 along the entire length of *CXCR4* in all three pairwise comparisons. On the other hand, peaks of dN/dS > 1 are observed in all three pairwise comparisons in *CD4*, two of which are significantly greater than 1 ($p < 0.001$ and $p < 0.004$; Figure 3-1C). The maximum dN/dS value found in each pairwise comparison is summarized in Figure 3-1D, where it can be seen that many genes have peaks of dN/dS >

1. The peaks mostly fall towards the beginning of the gene, the region encoding the D1 domain that interacts with HIV-1 (145). Sliding window analysis has an inherent multiple testing problem that is difficult to correct because of the non-independence of tests (windows overlap) (146). As an *ad hoc* method for eliminating some false positive signatures, we sought genes with regions of dN/dS significantly > 1 ($p < 0.05$) in at least two out of three different pairwise primate comparisons made. We find that 8 out of 40 genes analyzed meet these criteria (highlighted in gray in Figure 3-1D). Thus, we have identified preliminary signals of positive selection in eight candidate genes: *ANKRD30A, CD4, IBTK, MAP4, NUP98/96, NUP153, RANBP2,* and *WNK1*.

Now that we have potentially identified domains under positive selection, we wished to use a more sophisticated statistical test for positive selection, and to analyze dN/dS on a codon-by-codon basis. There are good methods for doing this, but such analyses require deep sequence sets for each gene to be analyzed (147). We next generated large primate datasets for each of these 8 candidate genes. Each gene was sequenced from 20 hominoid, Old World monkey, and New World monkey species (species shown with asterisks in Figure 3-2). For 5 of the 20 species, it was possible to acquire some gene sequences from available genome projects, and for the other 15 species cell lines were acquired and cDNA libraries constructed. Details of primate cell

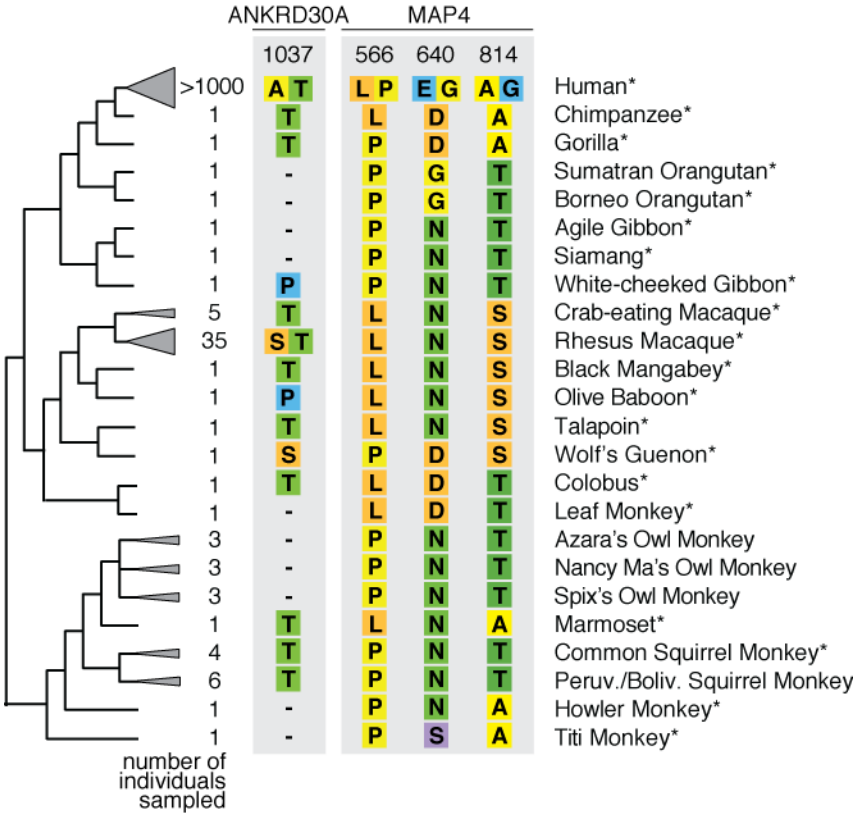


Figure 3-2: Genetic divergence and diversity in two candidate HIV-1 host factors. The amino acids at positions under positive selection are highly variable between species. Four positions are shown, one in *ANKRD30A* and three in *MAP4*. Numbers refer to amino acid coordinates. The colored boxes represent unique amino acids sampled at that position. The 20 species used in the positive selection analysis are indicated with an asterisk. All codons under positive selection in *ANKRD30A* and *MAP4* (not just these four) were re-sequenced from individuals representing small population sets of different non-human primate species (indicated with a triangle at the end of the branch on the primate cladogram). For humans, SNPs were identified in human SNP databases. Non-synonymous SNPs identified are indicated by 2 squares next to each other. A hyphen indicates lack of information, because *ANKDR30A* could not be sequenced from these species as described in the text.

lines, conditions for cell culture, mRNA extraction, cDNA library construction, PCR, and sequencing are given in the Materials and Methods section. High-quality Sanger sequencing was used. In all, 160 gene sequences for the 8 genes of interest were

generated. Two of these genes, *ANKRD30A* and *RANBP2*, were difficult to amplify from the available primate tissues, so sequencing was focused on a portion of these genes corresponding to the location of dN/dS > 1 peaks in the sliding window analysis, as indicated in Table 3-1. In the case of *ANKRD30A*, we were unable to amplify this gene from samples acquired for 7 species, so the final dataset for this gene consisted of only 13 species. In the case of *CD4*, sequences from 5 additional primate species, beyond the core set of 20 species, were available on Genbank. For each gene, the number of sequences in the final dataset, and the region of the gene analyzed, is summarized in Table 3-1.

The multiple sequence alignment generated for each gene was then analyzed for positive selection with PAML (93). Alignments were fit to a null model (Model M8a) where all codon positions are constrained to evolve with dN/dS ≤ 1, and a positive selection model (Model M8) where a dN/dS > 1 category of codons is allowed. A likelihood ratio test was then used to compare the positive selection model to the null model. The null model is rejected ($p<0.05$) in favor of the positive selection model in five of the eight genes analyzed: *ANKRD30A*, *CD4*, *MAP4*, *NUP153*, and *RANBP2* (Table 3-1). The null model could not be rejected in the case of *IBTK* (p=0.34), *NUP98/96* (p=0.46), and *WNK1* (p=0.79). For each of the 5 genes that passed this statistical test, the individual codons in the dN/dS > 1 class were identified (Table 3-1), and the residues corresponding to these codons are diagrammed in Figure 3-3. Figure 3-2 illustrates the amino acids encoded by four of these codons under positive selection, illustrating their

Table 3-1: PAML analysis of primate genes

Gene [a]	Number of species [a]	2Δl [b]	p-value [b]	Positive selection?	dN/dS [c]	% sites [c]	Codons with dN/dS>1 [d] * P>0.95 ** P>0.99
ANKRD30A aa 973-1335	13	36	p<0.001	**yes**	3.5	34%	A985, D986*, Q1015**, I1022*, V1027, N1032, T1037, C1049**, A1063*, S1100, Q1107*, I1169**, C1174**, E1191, L1211*, R1233**, M1239, F1268*, H1285, H1299*, E1302, L1309
CD4 aa 28-424	25	15	p<0.001	**yes**	2.2	18%	T15, T17*, Q20*, S23*, N32, I34, N39*, N52*, A55*, D88*, K206*, W214, R240*
IBTK aa 1-1353	20	0.9	p=0.340	no	na	na	na
MAP4 aa 10-1152	20	8.2	p<0.005	**yes**	5.5	1.0%	L566, T631, E640**, A814
NUP98/96 aa 80-1636	20	0.5	p=0.464	no	na	na	na
NUP153 aa 1-1475	20	8.9	p<0.003	**yes**	6.4	0.5%	I794, V1189**
RANBP2 aa 2002-3224	20	15	p<0.001	**yes**	3.7	2.2%	H2418**, A2724*, M2786*, M2813*, A2890*, T3153, Q3163, K3177*
WNK1 aa 742-2073	20	0.07	p=0.795	no	na	na	na

[a] Each dataset consists of gene orthologs from 20 primate species, with the exception of *ANKRD30A* due to inability to amplify this gene from 7 species, or *CD4* due to 5 additional primate sequences that were available in Genbank. The gene regions analyzed are indicated (by encoded amino acids, aa). Two of these genes, *ANKRD30A* and *RANBP2*, were difficult to amplify from the primate tissues that were available, so sequencing focused on the portion of these genes corresponding to the location of dN/dS > 1 peaks in the sliding window analysis.

[b] Twice the difference in the natural logs of the likelihoods (2Δl) of the two models (M8a-M8) being compared. The p-value indicates the confidence with which the null model (M8a) can be rejected in favor of the model of positive selection (M8).

[c] dN/dS value of the dN/dS>1 class of codons in M8, and the percent of codons falling in that class.

[d] Codons assigned to the dN/dS>1 class in M8 with a posterior probability of P>0.90 by Bayes empirical Bayes analysis. Codon coordinates and the encoded amino acid correspond to the human protein.

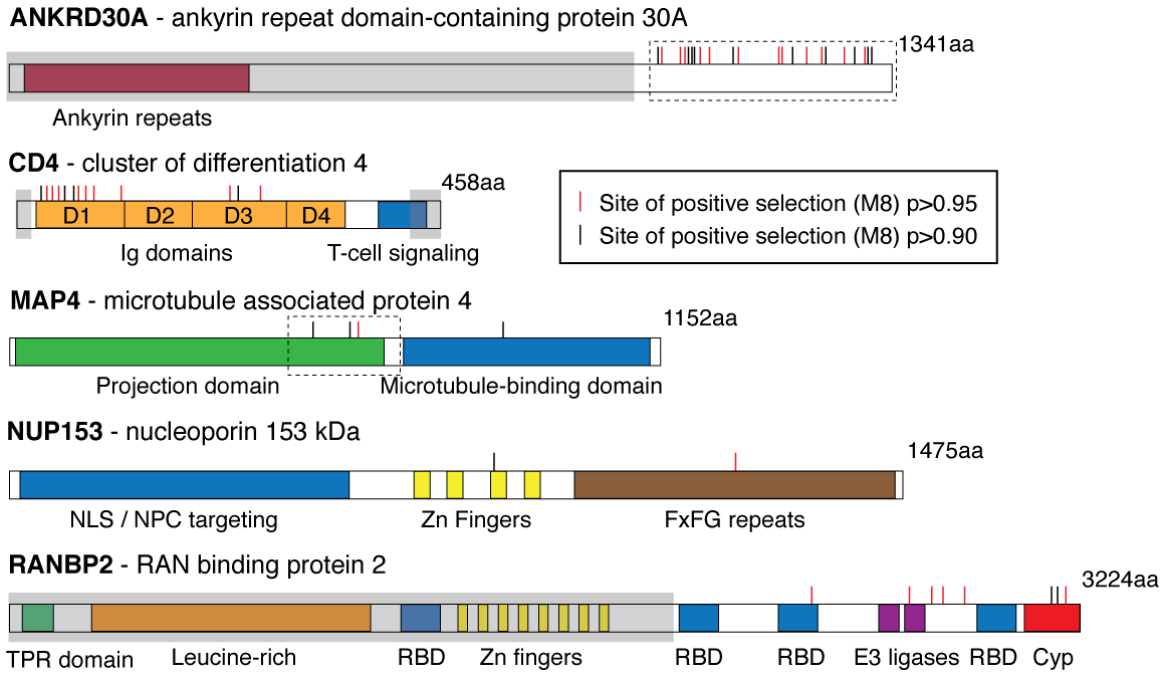


Figure 3-3: HIV-1 host factors exhibit signatures of positive selection. Domain diagrams are shown for the proteins encoded by each of the five genes evolving under positive selection. Residue positions experiencing positive selection are marked by red or black tick marks. Gene regions that were not sequenced are indicated by a transparent gray box. Dashed boxes indicate fragments used in infection assays in Figure 3-4. Diagrams are internally scaled.

dynamic nature over evolutionary time. In several instances we see "recycling" of amino acids at these positions, which has been observed before at positively selected sites (127).

These five genes are involved in various processes, with variable amounts known about their role in the HIV-1 lifecycle. CD4 is the main receptor for HIV-1 entry into cells (148) and has been previously documented as evolving under positive selection (99, 139). NUP153 and RANBP2, both associated with the nuclear pore (149), are known to be important for the trafficking of the HIV-1 pre-integration complex into the nucleus (101, 150-152). The positive selection of *RANBP2* has been previously noted in

mammals beyond primates (101). The identification of well-characterized host factors supports the idea that positive selection analysis can be a powerful addition to large-scale screening efforts. Two less characterized genes were also identified in our study: *ANKRD30A* and *MAP4*. While this work was in progress, the first report of the role of the microtubule-associated protein MAP4 in HIV-1 biology was published (153). Prior to this, little was known about the function of *ANKRD30A* other than the fact that it is upregulated in breast carcinomas (154). Nonetheless, *ANKRD30A* has the highest dN/dS value for any HIV-1 cofactor in this study, both by the metric shown in Figure 3-1B, and in the PAML analysis of individual codons (2Δl values in Table 3-1). No role for this gene in HIV-1 biology has yet been reported.

Characterization of ANKRD30A and MAP4

The novel candidate host factors *ANKRD30A* and *MAP4* were investigated further. A dominant negative assay similar to assays used to characterize the interaction of HIV-1 with the host factors NUP153, LEDGF, and CPSF6 was used (152, 155, 156). In this assay, expression of a protein fragment constituting the known or putative HIV-interaction domain of a host factor interferes with viral infection. This is because the fragment acts as a decoy that competes with the endogenous, full-length form of the same protein for interaction with virus proteins or complexes. This assay is particularly well-suited to this study, because positive selection is predicted to accumulate in portions of host factor genes corresponding to the virus-interacting domain of the encoded protein (127). Plasmids encoding truncated fragments of human ANKRD30A (amino acids 970-

1341) and MAP4 (amino acids 486-693) were generated that correspond to the regions where positively selected residues are found (expressed protein fragments are boxed in Figure 3-3). These FLAG-tagged constructs were stably expressed in CRFK cells (Figure 3-4A). Full-length MAP4 and ANKRD30A are both known to localize to the cytoplasm (153, 154, 157). The ANKRD30A fragment also localized to the cytoplasm, whereas the MAP4 fragment is distributed throughout the cell (Figure 3-4B). These cells were infected with retroviral vectors derived from HIV-1, feline immunodeficiency virus (FIV), and murine leukemia virus (MLV), all of which were pseudotyped with the vesicular stomatitis virus (VSV) G glycoprotein. Viral infection was measured by the expression of a green fluorescent protein (GFP) reporter gene. The expression of the ANKRD30A fragment reduced levels of HIV-1 and FIV infection by as much as 50%, but did not affect the integration of MLV (Figure 3-4E). The expression of the MAP4 fragment had no effect (Figure 3-4C-D), but mis-localization of the MAP4 fragment makes it difficult to draw any conclusions for this fragment. Nonetheless, this serves as a control for the specific effects of the ANKRD30A fragment on the inhibition of HIV-1.

ANKRD30A was previously identified in two independent RNA interference screens as important for HIV-1 replication, both of which used live HIV-1 and the infection of human cells (115, 118). Both of these screens also found this protein to be important for a step in the early part of the viral replication cycle, upstream of transcription and translation of viral genes and proteins (115, 118).

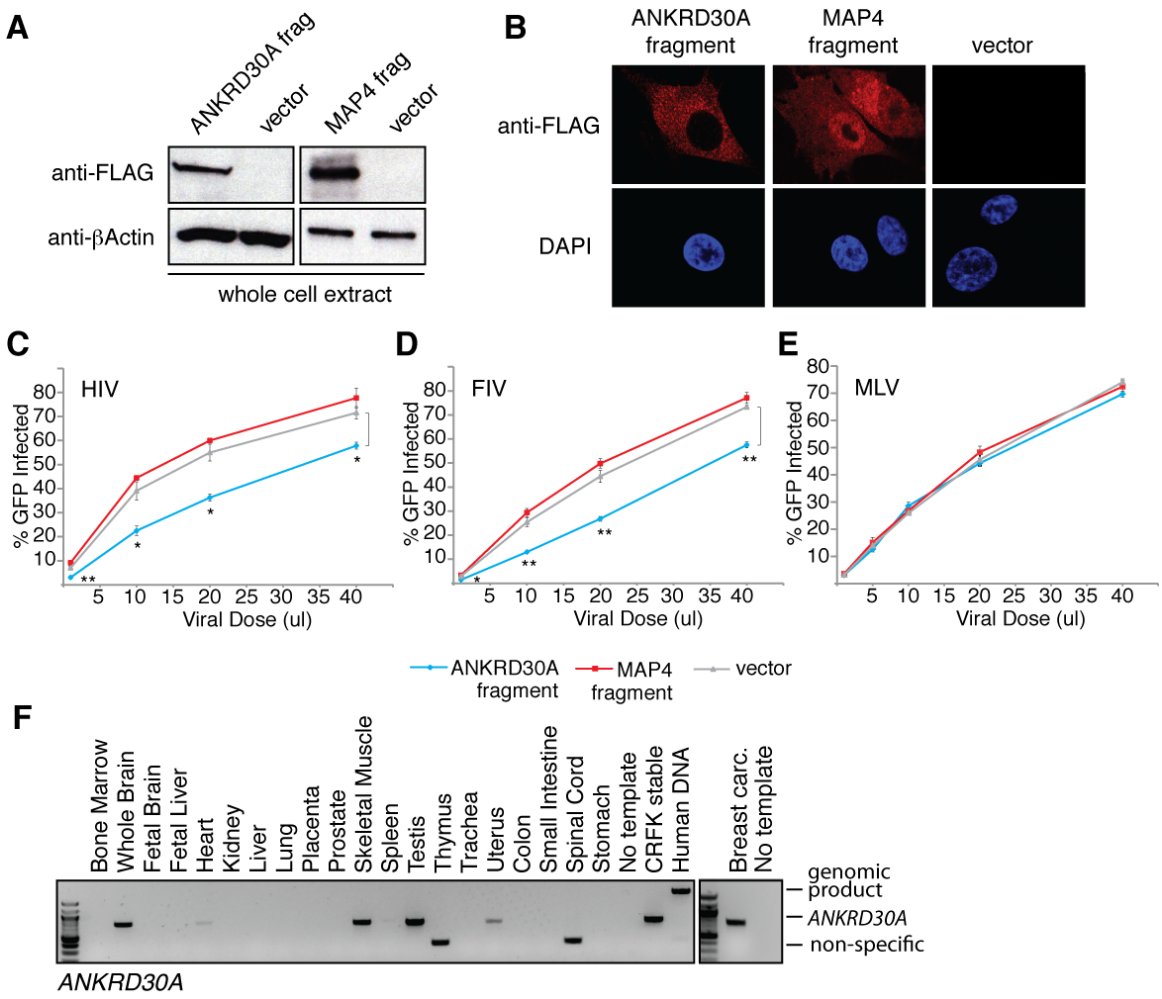


Figure 3-4: Characterization of ANKRD30A as a novel HIV-1 host factor.

Figure 3-4: Characterization of ANKRD30A as a novel HIV-1 host factor. (**A**) Western blot analysis of CRFK stable cell lines expressing FLAG-tagged fragments of ANKRD30A (amino acids 970-1341) or MAP4 (amino acids 486-693), or transduced with an empty vector (pLPCX). An actin antibody serves as a loading control. (**B**) Confocal microscopy was performed for each of the three cell lines. Cells were stained with a primary FLAG-specific antibody and with DAPI, and then visualized using a secondary Alexa Fluor 594 antibody. Images shown are representative of each stable cell line. (**C-E**) Single-cycle infection assay using VSV-G pseudotyped retroviruses. For each cell line, dose curves were generated in triplicate and standard deviations are indicated with error bars. Significant differences relative to the empty vector control were calculated using a two-tailed Student's t-test. (*) $P<0.01$ (**) $P<0.001$ (**F**) A human tissue panel was probed for *ANKDR30A* expression using a primer set that spans an intron. As a positive control, the CRFK cells stably expressing the dominant negative ANKRD30A fragment were included. Human DNA was also used as a control, to show the product size obtained from genomic template (with the intron). Because *ANKRD30A* was originally identified as a gene associated with breast cancers, its expression was also demonstrated in a breast carcinoma cell line.

These data place the timing of interaction between HIV-1 and ANKRD30A in a slightly narrower window, between cellular entry and nuclear entry. It also suggests that this interaction is direct, and that the final third of the ANKRD30A protein constitutes the virus interaction domain. The most exposed HIV-1 protein at this stage of the viral lifecycle is capsid. To investigate a possible direct interaction between capsid and this fragment of ANKRD30A, this fragment of ANKRD30A was fused to the TRIM portion of the owl monkey TRIM-Cyp restriction factor (see (101)(158) for a description of this assay). Viral restriction indicative of an interaction between capsid and TRIM-ANKRD30A was not observed, but very low expression of these TRIM fusion proteins was noted, suggesting that they may be misfolded or unstable (data not shown). ANKRD30A is expressed in white blood cells (GenBank record BF171216.1), and these data show expression in brain, heart, muscle, testis, and uterus (Figure 3-4F). More extensive work will be needed to characterize the interaction of ANKRD30A with HIV-1 and to investigate a possible role of ANKRD30B in HIV-1 biology. This could be quite important because, in addition to having the strongest signature of positive selection in this study, ANKRD30A is one of the few candidate HIV-1 host factors that has been categorized as a "druggable" target, based on properties that is shares with proteins that have been already successfully targeted by drugs (22, 159).

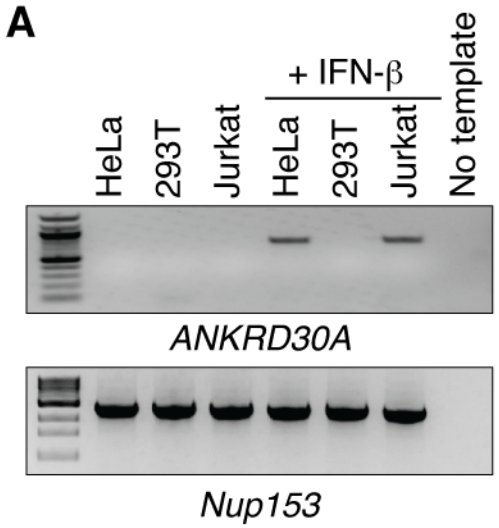
In addition to species-specific genetic differences, single nucleotide polymorphisms (SNPs) circulating within primate species can dramatically affect interactions between host restriction factor proteins and retroviruses (28, 122, 160, 161). For instance, SNPs found in the *TRIM5* restriction factor locus of rhesus macaques can

dramatically affect the ability of the encoded TRIM5α protein to restrict different HIV-1 and simian immunodeficiency virus (SIV) strains (16). Codons targeted by positive selection in *MAP4* and *ANKRD30A* (Table 3-1 and Figure 3-3) were sequenced from small populations of select primate species. This population survey focused on three owl monkey species (Azara's, Nancy Ma's, and Spix's owl monkeys), three squirrel monkey species/sub-species (common, Peruvian, and Bolivian squirrel monkeys), and two macaque species (crab-eating, also known as cynomolgus, and rhesus macaques). We also utilized human SNP databases to look for non-synonymous human SNPs that coincide with codons under positive selection. *MAP4* bears human SNPs at three out of four sites of positive selection, but no polymorphisms at these sites in the macaque, squirrel monkey, or owl monkey populations that we investigated (Figure 3-2). In *ANKRD30A*, SNPs were found at one site of positive selection (residue 1037) in both human and rhesus macaque populations. *ANDRD30A* from owl monkeys was unable to be sequenced. Unlike all other genes analyzed, the inability to sequence *ANKRD30A* from some species was a consistent problem in this study, suggesting that this gene is being lost, duplicated, or rearranged in different species. More extensive work will be needed to characterize the interaction of MAP4 and ANKRD30A with HIV-1, and to determine the significance of species-specific and population-level genetic divergence in these genes.

Inconsistent identification of *ANKRD30A* in different genetic screens

Interestingly, *ANKRD30A* was identified as critical for HIV-1 replication in some but not all genome-wide RNA interference screens. It was identified in two screens conducted in HeLa cells (115, 118), but not in similar screens performed in 293T or Jurkat cells (116, 117). ANKRD30A was also not identified in a proteomic screen for HIV-interacting proteins that was conducted in both 293 and Jurkat cells (162). To investigate this discrepancy, the expression of *ANKRD30A* in HeLa, 293T, and Jurkat cells was tested. Interestingly, *ANKRD30A* was not expressed in any of these cell lines (Figure 3-5A), making it hard to understand why this gene would have been identified in any screens. The transfection of short interfering RNAs (siRNAs), which was performed in several of these screens (Figure 3-5B), is known to induce the interferon response (163), and siRNAs may be detected by pattern recognition receptors in the cytoplasm. We considered the possibility that *ANKRD30A* expression was induced under some experimental conditions by siRNA transfection and interferon induction, and then subsequently knocked down when on-target siRNAs were used. To explore this idea, we treated HeLa, 293T, and Jurkat cells with interferon-β and found that expression of *ANKRD30A* is readily detected in HeLa and Jurkat cells, but not in 293T cells (Figure 3-5A).

These expression patterns may explain why *ANKRD30A* was found in some genome-wide screens and not others (summarized in Fig. 3-5B). It was identified only in screens that combined HeLa cells with siRNA transfection (115, 118). Jurkat cells are not easily transfectable, so transduction was always used with these cells. In transduction, the



B

Screen	Identified ANKRD30A?	Cells Used	Expression Method
Brass	Yes	HeLa	siRNA transfection
Zhou	Yes	HeLa	siRNA transfection
Konig	No	293T	siRNA transfection
Yeung	No	Jurkat	shRNA transduction
Jager	No	293	HIV gene transfection
		Jurkat	HIV gene tranduction

Figure 3-5: *ANKRD30A* expression under conditions used in high-throughput genetic screens. (**A**) Expression of *ANKRD30A* was assessed in various laboratory cell lines. Cells were grown both without and with stimulation by interferon-β, and mRNA was harvested. Expressed and spliced transcripts were detected by RT-PCR. Primers were also designed to amplify spliced transcripts of *NUP153*, and were used as a control for the amount of input material in each reaction. (**B**) A summary of genome-wide screens for HIV-1 host factors, along with relevant details of the methods used.

retroviral capsid core would be predicted to shield nucleic acids from detection by pattern recognition receptors. This is a demonstration of how genome-wide screens may be sensitive to the particular cell type used, and why different studies may not agree because of differences in the experimental design.

DISCUSSION

Here five human genes that both promote HIV-1 replication and contain signatures of positive selection were identified. Although the positive selection of restriction factor genes is well documented, it is becoming clear that arms races may also be shaping host genes that facilitate pathogen life cycles, with preliminary reports coming

from cell surface receptors (25, 89, 90, 137) and DNA repair machinery (92, 138). In the current study, 5 out of 40 housekeeping genes investigated (13%) were found to have experienced strong positive selection. We conclude that HIV-1 host factor genes can be caught in evolutionary arms races just like restriction factor genes. However, in this case, collateral damage to cellular housekeeping functions must be carefully controlled as the evolutionary battle with viruses plays out. The identification of three well-known HIV-1 host factors in our screen supports the idea that intersecting gene lists from HIV-1 host factor screens with evolutionary analysis for positive selection will provide a meaningful way to prioritize certain candidate host factors over others.

ANKRD30A was identified as a novel HIV-1 host factor of interest. This gene has the strongest signature of positive selection observed in this study. A truncated form of ANKRD30A inhibited HIV-1 and FIV, but not MLV, in single-cycle infection assays. The protein region corresponding to this dominant-negative fragment contains many residues targeted by positive selection (Figure 3-3), as well as SNPs in both human and primate populations (Figure 3-2). Together, these data are consistent with the C-terminal portion of ANKRD30A making direct physical contact with an unknown HIV-1 protein or complex. This gene is expressed in $CD4^+$ Jurkat T-cells, the target tissue for HIV-1 replication, at least under conditions of interferon stimulation. The interferon induction of this gene is intriguing, as one would not expect a host factor to be upregulated during an immune response. However, this gene is also expressed in a handful of normal tissues (Figure 3-4F). *ANKRD30A* has been identified as copy number variable in the human

population (164), and gene dosage of an HIV-1 host factor could contribute to variable susceptibility of disease progression.

With the advent of high-throughput genetic and proteomic screening in the field of HIV-1 research, there is now an opportunity to utilize techniques from the field of systems biology so that the results of multiple high-throughput studies may be integrated (165). The field of systems biology provides a roadmap for finding "signal in the noise" generated by these screens, predominantly by looking for associations that come up repeatedly in independent types of studies and assays. It is important that all types of screens continue to be performed, as each has its strengths and weaknesses (106). Yeast two-hybrid screens capture direct human-HIV physical interactions, pull-down/mass-spectrometry based screens capture interactions that occur in the context of larger protein complexes, and genetic screens capture genes with even indirect effects on viral replication. Genome-wide association studies yield different information altogether, in this case revealing information only on genes where polymorphism between humans results in differential susceptibilities (76, 166, 167). Positive selection analysis is yet another type of screen that can and should be added into this systems biology approach, also having its own strengths and weaknesses. A particular strength is that this analysis tends to identify genes that encode proteins that interact with HIV-1 directly (127).

Several limitations of positive selection analysis must also be noted. First, the positive selection of human and primate genes can be driven by processes other than host-virus arms races. For instance, genes involved in sexual selection can experience positive selection (168). In the case of the present study, we began not with a random set

of genes, but with a set that had scored in assays for HIV-1 replication, making retroviruses the most likely drivers of this evolutionary signature. Second, the evolutionary signature in a gene can be a composite of multiple selective forces, potentially involving host-virus arms races taking place with more than one virus family (137). Another limitation is that some host genes involved in retroviral lifecycles will not experience positive selection. Indeed, *CXCR4*, *LEDGF* and *CPSF6* are examples of human genes that are well-known to be involved in HIV-1 replication, and which encode proteins that directly interact with retroviruses, but that are extremely conserved in protein sequence (158, 169). From the sliding window analysis of dN/dS (Figure 3-1D), another gene recently characterized important to the HIV-1 lifecycle, *TNPO3* (170, 171), also appears to be highly conserved. These genes apparently lack the evolutionary flexibility to engage in evolutionary arms races. In some cases, genes are so essential that no mutation exists that could offer an advantage in the face of viral challenge without serious collateral damage to the normal cellular function of this gene. Other host genes will be able to produce mutations that dissociate host functions from viral interactions, but these mutations may act in a recessive fashion with regard to viral susceptibility. Such mutations will not experience selection in the context of heterozygous hosts, making it unlikely that they will become common enough for homozygotes to arise (where selection would start acting). One major difference between the evolution of restriction factors and host factors is that mutant alleles of restriction factors act in a dominant or semi-dominant fashion, whereas alleles of host factors are predicted to act in recessive or semi-dominant fashion (127).

In conclusion, positive selection analysis can be used to strengthen high-throughput and systematic studies of host-virus interactions. Although this analysis has certain limitations, its strength and value increases when combined with other types of virological data, as demonstrated here.

Chapter 4

Identification of the first nonhuman primate CD4 receptor compatible with primary HIV-1 isolates

Because HIV-1 does not replicate in most mammals, studies of viral pathogenesis employ modified viruses known as SHIVs. SHIVs encode an envelope protein (Env) that can use macaque CD4 as an entry receptor. However, these Envs are not representative of the major circulating variants of HIV-1. Therefore, the major antigen on HIV-1 relevant to vaccine studies, Env, is altered in the course of creating the animal model used to study vaccines. Animal models for SHIV strains bearing clinically-relevant Envs are lacking. Here, *CD4* genetic diversity within primate populations was sampled. Spix's owl monkey (*Aotus vociferans*) individuals were found to carry *CD4* alleles compatible with HIV-1 Envs isolated from early human infections, and representing all of the major HIV-1 group M clades (A, B, C, and D). Based on these results, a strategy for building improved HIV-1 model organisms is propsed. If permissive alleles at key loci like *CD4, CCR5,* restriction factor genes, and *MHC* could be identified within a single primate species, they could be combined into a single animal through mating.

INTRODUCTION

Humans, chimpanzees and white-handed gibbons are the only mammals that are known to support HIV-1 replication, but the latter two primates are endangered and only rarely develop immunodeficiencies upon infection (172). Cells from other nonhuman primate species are resistant to HIV-1 infection even in laboratory cell culture, in most cases due to restriction factors that they encode (11, 173). However, entry into the cell is also a major barrier to HIV-1 infection in many nonhuman primate species. For example, HIV-1 variants derived directly from humans at early stages of infection, which are most relevant to the HIV-1 pandemic, have only been shown to be compatible with human CD4 (174), whereas lab-adapted or chronic-stage isolates of HIV-1 can use the CD4 receptor encoded by multiple nonhuman primate species (175, 176). In the laboratory, mutations can be introduced into virtually any cellular receptor or restriction factor to make it compatible with HIV-1 replication, but using these modified genes to create transgenic laboratory animals, particularly monkeys, has been a consistent roadblock in translating these findings.

In the absence of an animal model for HIV-1, the main strategy thus far has been to modify the virus itself, thereby making it capable of replicating in some macaque species and even inducing an AIDS-like disease (172, 177). This has given rise to a number of SHIV strains (simian/human immunodeficiency virus hybrids) that replicate well in macaques, but only after serial passage and adaptation in this species. The process of adapting SHIVs to replicate in macaques leads to changes in the viral envelope glycoprotein (Env) (178), as well as the ability of HIV-1-specific antibodies to recognize

Env (Boyd, Peterson, Haggarty, Jordan, Hogan, Goo, Hoxie and Overbaugh; submitted). Because Env is the major antigen relevant for vaccine studies, this potentially compromises the macaque model for some applications and could contribute to the failure of vaccine studies to translate into humans. In addition, there are currently no SHIVs encoding Envs derived from HIV-1 variants isolated soon after sexual transmission, which is the major mode by which HIV-1 is spread. Such SHIVs do not replicate in nonhuman primates, at least in part due to incompatibilities with the CD4 receptor, an issue that has only recently been appreciated (174, 179). Instead, most SHIVs bear Env from lab-adapted HIV-1 strains, typically isolated during chronic stages of infection. Finally, the majority of SHIVs represent subtype B Env sequences, and few are representative of subtypes A, C, and D, which are the most prevalent types in sub-Saharan Africa.

There remains an urgent need to identify a nonhuman primate species that will support the replication of SHIVs bearing Env from clinically-relevant human isolates. In this pursuit, the population genetics of nonhuman primate species has largely been untapped. Like in humans, significant genetic polymorphism exists in nonhuman primate populations. Here, we explore the possibility that natural variation already exists within primate species that could be harnessed for the development of an animal model. In this study, small populations were analyzed, representing 11 different nonhuman primate species, and individual animals were evaluated for *CD4* polymorphisms. Multiple *CD4* alleles were identified in one New World monkey species, the Spix's owl monkey (*Aotus vociferans*), which encode CD4s that support entry mediated by Envs from all of the

major clades of HIV-1 group M. This very limited survey of primate populations is a powerful demonstration of how large-scale bioprospecting in nonhuman primate populations could reveal naturally-occurring alleles compatible with HIV-1 infection. If permissive alleles at key loci like *CD4, CCR5,* restriction factor genes, and *MHC* could be identified within a single primate species, they could be combined into a single animal. Such an approach would require no special transgenic methods, but rather simple breeding techniques dating back thousands of years.

MATERIALS AND METHODS

Primate samples

Nonhuman primate samples were acquired from the M.D. Anderson Cancer Center's Michale E. Keeling Center for Comparative Medicine and Research (KCCMR) in Bastrop, TX, or from the New England Primate Research Center in Southborough, MA. For each individual, 2.5 mL of blood was collected in PaxGene Blood RNA Tubes (BD, #762165)**.** Alternately, B-cell lines were expanded in suspension culture in RPMI, 20% FBS, Pen/Strep, L-glutamine, HEPES, and AZT. Genomic DNA and RNA from blood and cell lines were isolated using the PaxGene miRNA kit (Qiagen, #763134) and/or the Qiagen All Prep DNA/RNA mini kit (Qiagen, #80204). cDNA libraries were generated using oligo(dT) primers with isolated primate RNA and the Superscript III First-strand Synthesis System (Invitrogen, #18080-051). PCR was performed from gDNA or cDNA with PCR SuperMix High Fidelity (Invitrogen, #10790-020) or Phusion

High Fidelity PCR Master Mix (NEB, #F-531S). All newly characterized SNPs were verified with independent PCR and sequencing reactions.

Expression constructs.

Human *CD4* was amplified from RNA isolated from Jurkat T cells. Owl monkey *CD4* was amplified from RNA isolated from blood samples as described above. Each *CD4* was sub-cloned into the pCR8 Gateway entry vector using TA cloning (Invitrogen, #K2500-20). An LR Clonase II reaction (Invitrogen, #11791-100) was used to move these constructs into a Gateway-converted pLPCX retroviral packaging vector (Clontech, # 631511). The expression plasmid encoding rhesus macaque CD4 was described previously (180).

Envelope clones

The following envelope clones from early HIV-1 infections were used: Q461e2 (181), QH343.21 (182), WITO4160.33 and TRO.11 (183), ZM53M.PB12 and ZM197M.PB7 (184), QA013.70I and QB857.110I (182). As a control, two subtype B HIV-1 *env* clones (BaL.01 and SF162) representing variants known to infect macaque cells (174) were also used. GFP reporter pseudoviruses were generated in HEK293T cells by cotransfecting 667 ng of Q23Δ*env*-GFP (179) and 333 ng of the HIV-1 *env* clone of interest using Fugene 6 (Roche) transfection reagent at a ratio of 3 μl Fugene 6 to 1 μg DNA following the manufacturer's protocol.

Generation of stable cell lines

HEK293T and Cf2Th/syn CCR5 (185) cells were cultured in Dulbecco's modified Eagle's medium (Invitrogen) with 10% FBS (Gibco), 2 mM L-Glutamine (Gibco), and 1% antibiotic (Gibco) (complete medium) at 37°C and 5% CO_2. Cf2Th/syn CCR5 cells, which are canine thymocytes engineered to express human CCR5, were further supplemented with 400 μg/ml of Geneticin (Gibco) to maintain CCR5 expression. For generation of CD4-expressing cell lines, retroviral virus-like particles (VLPs) were generated in HEK293T cells by cotransfecting pLPCX (retroviral vector encoding the CD4 of interest), pJK3 (MLV-based packaging plasmid), and pMD.G (VSV-G envelope plasmid) at a ratio of 1:1:0.5 using Fugene 6 (Roche) transfection reagent following the manufacturer's protocol. Forty-eight hours post-transfection, the supernatant containing VLPs were collected, filtered through 0.22 mm filters, and concentrated using Amicon Ultracel 100K filters (Millipore). The concentrated VLPs (~200 μl) were used immediately to transduce Cf2Th/syn CCR5 cells that had been plated 24 h prior at a density of 10^5 cells/well in a 6-well plate in 2 mL of drug-free complete medium. The cells were transduced in the presence of 10 μg/ml of DEAE-dextran by spinoculation at 1200 *g* for 90 min. The following day, cells were split and transferred into new T75 flasks in 10 ml of drug-free complete medium and cultured for 48 h. The cells were then passaged and maintained in complete medium supplemented with 400 μg/ml of Geneticin (to maintain CCR5 expression) and 2 μg/ml of Puromycin (to select for CD4 expression). The transduced cells with high levels of CD4 expression were obtained by sorting the cells on FACSAria II cell sorter using an allophycocyanin-conjugated CD4 monoclonal

antibody (BD Biosciences, #551980) as described previously (174). Cf2Th/syn CCR5 cells stably expressing Rhesus CD4 have been described previously (174).

CD4 infectivity assay

Cf2Th/syn CCR5 cells stably expressing CD4 (2.5 x 10^4 cells/well in a 12-well plate in 1 ml of drug-free complete medium) were seeded 24 h prior to infection. The cells were infected with HIV-1 pseudoviruses in duplicate wells at an MOI of 0.5 in the presence of 10 μg/ml of DEAE-dextran by spinoculation at 1200 *g* for 90 min. After 72 h, the cells were washed once with 200 μl 1x PBS, harvested using 200 μl of 0.05% Trypsin-EDTA (Gibco), and fixed in 200 μl of 2% paraformaldehyde. The fixed cells were washed twice with 500 μl FACS buffer (1x PBS buffer containing 1% FBS and 1 mM EDTA). The cells were resuspended in 400 μl of FACS buffer, filtered through a 35 μm pore size nylon mesh cap (BD Falcon) and analyzed for GFP expression on a BD FACSCanto II flow cytometer. The data from ~10^4 cells were analyzed using FlowJo version 9.7.5.

Evolutionary analysis

An alignment of primate *CD4* D1 domain was analyzed using the codeml program contained in PAML 4 (100). The free-ratio model was used to estimate dN/dS and the number of non-synonymous and synonymous substitutions that occurred along each branch.

RESULTS

Polymorphism in the primate *CD4* gene

To detect single nucleotide polymorphisms (SNPs), *CD4* from small populations of select nonhuman primate species (bold type in Figure 4-1) were sequenced. Although in practice macaques have been used as the main animal model, there has also been interest in developing New World monkeys as a model system (176, 186). For this reason, this population survey focused on two macaque species (cynomolgus and rhesus macaques) and five different species of New World monkeys: three owl monkey species (Azara's, Nancy Ma's, and Spix's owl monkeys) and two squirrel monkey species (common squirrel monkeys, and the related Peruvian and Bolivian squirrel monkey sub-species). For this study, whole blood or B-cell lines were obtained from primates housed at two different primate research centers. The number of individual monkeys sampled from each species ranges between 3 and 35, as indicated in Figure 4-1. Human SNP databases and previous reports on *CD4* diversity in chimpanzees, three different species of African green monkeys (grivets, tantalus monkeys, and sabaeus monkeys), and squirrel monkeys (187-189) were also used.

CD4 interacts with HIV-1 Env through its D1 protein domain (145). Even in the relatively small populations surveyed, 6 of the 12 primate species exhibited

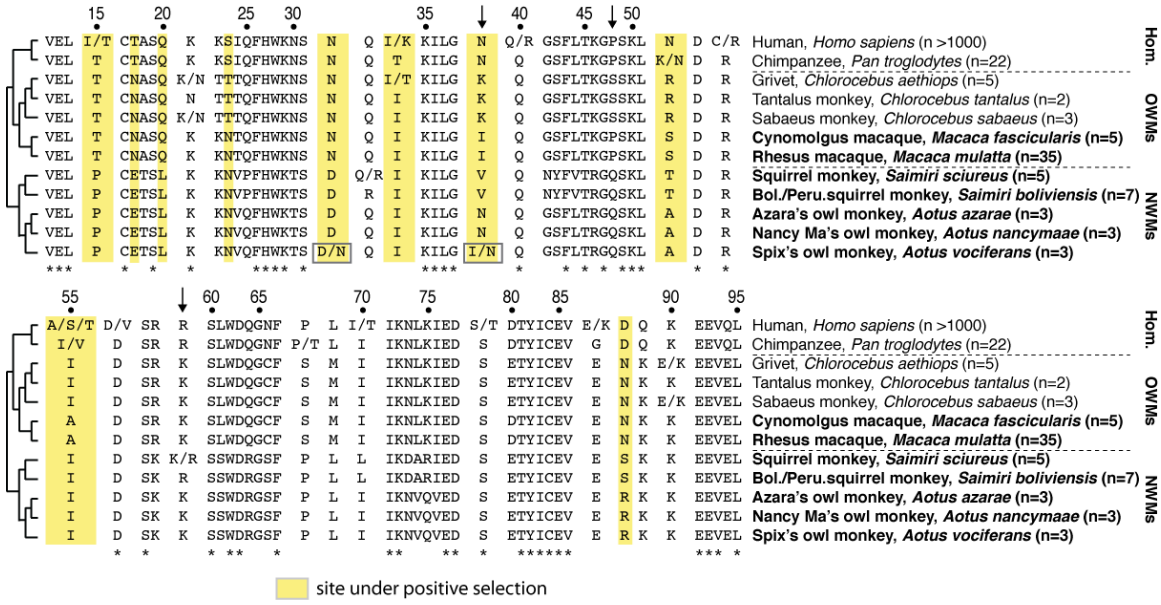


Figure 4-1: Non-synonymous SNPs in the portion of *CD4* encoding the D1 domain. An amino acid alignment of the D1 domain from species included in the population study. Non-synonymous SNPs identified are indicated with slashes, where the two alternately encoded amino acids are given on either side. Dotted lines separate the major clades of simian primates (Hom. = hominoids, OWMs = Old World monkeys, NWMs = New World monkeys). The number of individuals analyzed from each species is shown in parentheses adjacent to the species name. Amino acid positions highlighted in yellow were previously identified to be evolving under positive selection (27). Numbering along the top is relative to the mature CD4 protein, after cleavage of the 25 amino acid N-terminal signal peptide. Species shown in bold indicate populations that were sequenced in this study. Arrows indicate three sites that have been shown to affect HIV-1 entry.

non-synonymous polymorphism in this domain (slashed positions in Figure 4-1). A total of 17 sites in the D1 domain contain non-synonymous SNPs. Species-specific (as opposed to intra-specific) variation at three sites in the D1 domain, N39, P48, and R59, has been shown to alter interactions with HIV-1 (arrows in Figure 4-1) (174, 189). For instance, a single amino acid difference at position 39 of CD4 between human (asparagine) and pig-tailed macaque (*Macaca nemestrina,* isoleucine) accounts for the species-specific differences in the ability of these CD4s to function as receptors for HIV-

1 (174). Of these three positions, only N39 is under positive selection (27) and is therefore highly variable between species. At this position, various nonhuman primate species encode asparagine (N), isoleucine (I), lysine (K), or valine (V) (Figure 4-1). This position is also polymorphic in the Spix's owl monkey species, where certain alleles encode the amino acid found at position 39 in humans (asparagine; N) and other alleles encode the amino acid found in macaque species (isoleucine; I) (Figure 4-1). On the basis of these findings, we hypothesized that there is functional variation in the CD4 proteins encoded by different Spix's owl monkey individuals.

Spix's owl monkey *CD4* alleles are compatible with entry by HIV-1 isolates from early human infections

To test the ability of Spix's owl monkey CD4 to support HIV-1 infection, we cloned all four *CD4* alleles from two individuals that encode non-synonymous SNPs at residue 39, or a neighboring site under positive selection, residue 32 (both boxed in Figure 4-1). These alleles encode proteins that are identical at every other amino acid position with one exception: the Spix 2 and Spix 3 (Figure 4-2A) alleles encode identical proteins but differ by a T275A SNP outside of the D1 domain (not shown in alignment). Cell lines stably expressing these *CD4* alleles were generated in Cf2Th/syn CCR5 cells, which express human CCR5. Cells expressing human or rhesus CD4 were also generated as controls. Similar expression levels of CD4 were observed across these cell lines (Figure 4-2B).

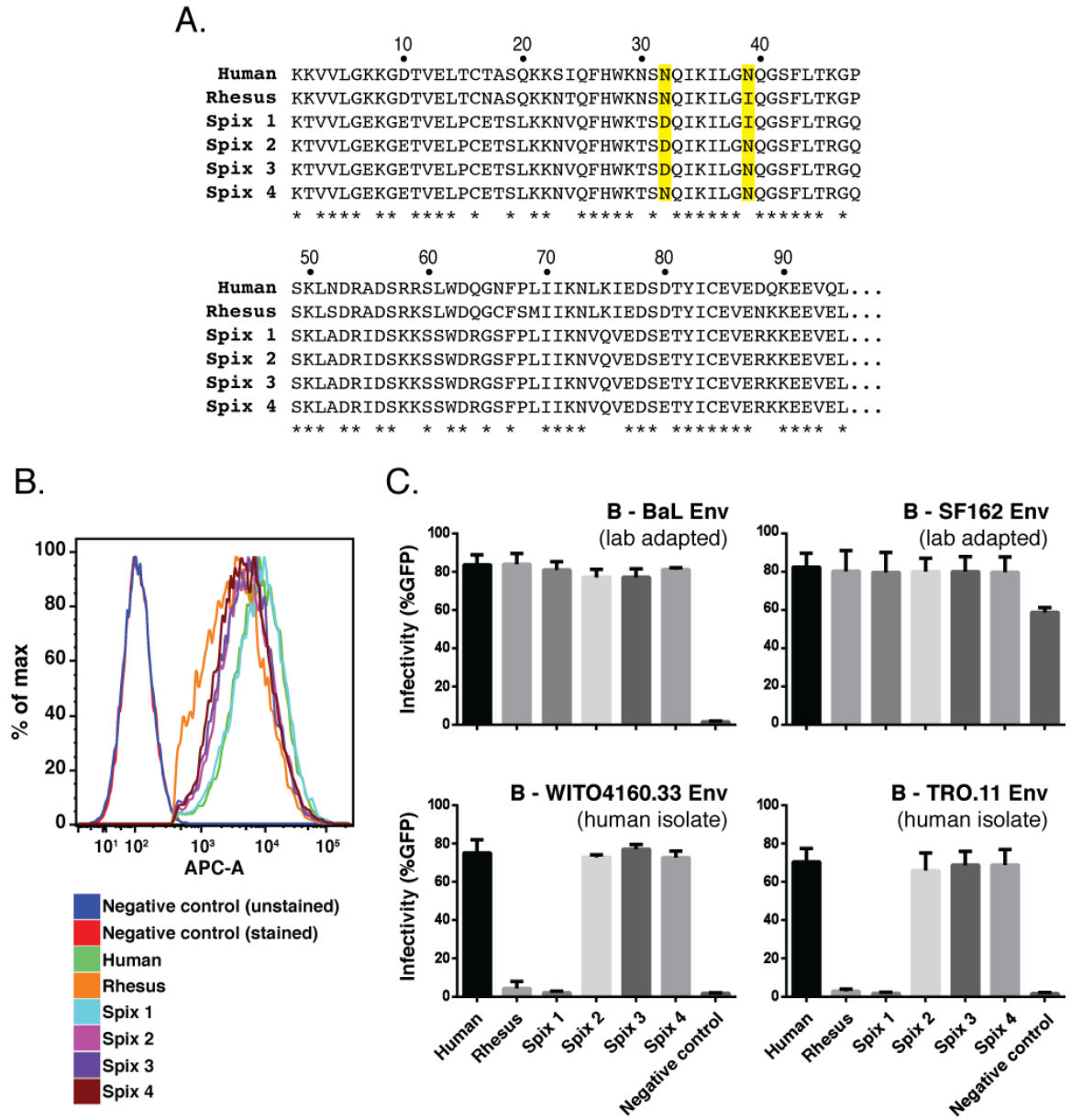


Figure 4-2: Spix's owl monkey *CD4* alleles encode receptors permissive for entry by primary HIV-1 isolates.

Figure 4-2: Spix's owl monkey *CD4* alleles encode receptors permissive for entry by primary HIV-1 isolates. **(A)** Amino acid alignment of the D1 domain from six CD4 receptors tested for virus entry. Sites that are polymorphic in Spix's owl monkeys are highlighted in yellow. **(B)** Expression levels of CD4s stably introduced into Cf2Th/syn CCR5 cells as measured by flow cytometry using an APC-conjugated CD4-specific antibody. **(C)** Cell lines expressing different *CD4* alleles (denoted along x-axis) were infected with the indicated Env-pseudotyped virions. Env variants are labeled such that the first letter represents the subtype, followed by the strain name. Infection is indicated as percent GFP-positive cells measured by flow cytometry 72 hours post infection. All pseudotyped virions were generated using the subtype A-derived Q23DEnvGFP proviral clone. Error bars represent the standard deviation of the mean from three independent experiments conducted in duplicate.

Virions bearing Envs from two lab-adapted subtype B variants (BaL and SF162; (174)) gained entry through all CD4s tested, including those encoded by the 4 Spix's owl monkey alleles (Figure 4-2C). The env genes from subtype B viruses isolated from early-stage human infections, including one that was not passaged prior to cloning (WITO4160.33) and one that was cloned after low passage (TRO.11) (183) were tested. As observed previously (174), human CD4, but not rhesus macaque CD4, supports entry of these viruses (Figure 4-2C). This illustrates the key obstacle that CD4 presents in macaques, requiring that adapted Envs are used. Surprisingly, three out of the four Spix's owl monkey CD4 receptors tested were permissive to HIV-1 pseudotyped with these Envs, at a level similar to human CD4 (Figure 4-2C). The three Spix's owl monkey CD4 receptors that support infection (Spix 2, 3, and 4) all have an asparagine (N) at position 39, whereas Spix 1 encodes an isoleucine (I) at this position (Figure 4-2A). This amino acid mutation is solely responsible for this phenotype, because Spix 1 and Spix 3 encode identical proteins except at position 39.

Spix's owl monkey *CD4* alleles are permissive to major HIV-1 subtypes

To further investigate the permissiveness of these Spix's owl monkey *CD4* alleles for HIV-1 entry, the CD4-expressing cell lines were challenged with a panel of viruses pseudotyped with Envs from the major clades of HIV-1 group M circulating globally

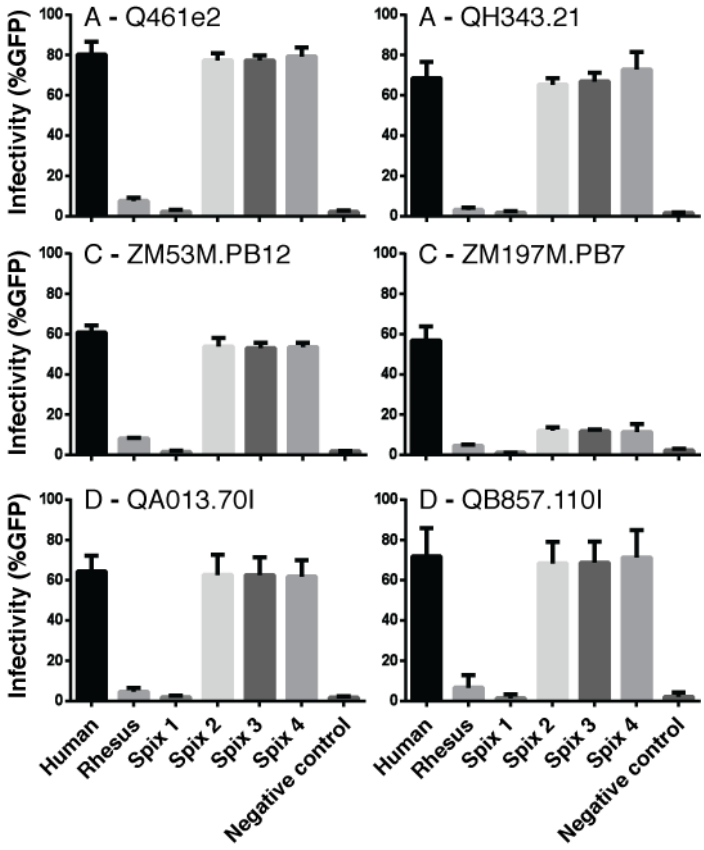


Figure 4-3: Spix's owl monkey CD4 receptors are permissive for entry by the major clades of HIV-1 group M. Cf2Th/syn CCR5 cells expressing various *CD4* alleles (denoted along x-axis) were infected with the indicated Env-pseudotyped virions. Env variants are labeled such that the first letter represents the subtype, followed by the strain name. Infection is measured by flow cytometry as percentage GFP-positive cells 72 hours post infection. All pseudotyped virions were generated in the Q23DEnvGFP background. Error bars represent the standard deviation of the mean from three independent experiments conducted in duplicate.

(Figure 4-3). This panel includes R5-tropic Envs from two subtype A variants (Q461e2, QH343.21), two subype C variants (ZM53M.PB12 and ZM197M.PB7), and two subtype D variants (QA013.70I and QB857.110I). All of these Envs are derived from HIV-1

isolated in early phases after sexual transmission and, importantly, are representative of the most prevalent circulating variants of HIV-1.

Human CD4 facilitates entry by all of these viruses (Figure 4-3). In contrast, the rhesus macaque CD4 receptor does not facilitate entry by any of them, again dramatically illustrating the major hurdle in generating SHIVs that are both replication-competent in macaques and relevant for human vaccine design. The three Spix's owl monkey *CD4* alleles that encode an asparagine (N) at position 39 (Spix 2, 3, and 4) encode receptors that facilitate entry by 5 of the 6 Envs tested. These alleles support entry by at least one Env from each subtype tested (A, B, C, or D). Therefore, the CD4 encoded in the genomes of some Spix's owl monkeys are broadly permissive to entry by HIV-1 group M subtypes. In this limited survey of three individual Spix's owl monkeys, two of the individuals are homozygous for permissive alleles.

Interestingly, not all subtype C Envs can enter through these Spix's owl monkey CD4s (Figure 4-3). Subtype C Env ZM197M.PB7 is unable to use any Spix's owl monkey CD4, whereas subtype C Env ZM53M.PB12 readily enters through CD4 encoded by three of the Spix's owl monkey alleles. This result illustrates how genetic diversity in both the host and the virus can impact whether the CD4 receptor can mediate entry.

A host-virus evolutionary arms race involving primate CD4

Why does the ability of CD4 to act as an HIV-1 receptor vary so much between and within species? To answer this question, we considered the evolutionary pressures

shaping the *CD4* gene. We and others previously demonstrated that *CD4* is evolving under positive natural selection in primates (27, 139). When hosts and viruses co-evolve over long periods of time, both will experience positive selection for new allelic forms of genes that regulate their interaction with one another (126, 127). For instance, primate genomes will be selected for new allelic forms of *CD4* that limit retroviral entry. Retroviruses will, in turn, be selected for new allelic forms of *env* that permit entry using new forms of CD4. This dynamic results in the accelerated fixation of mutations at the binding interface between host and virus proteins (CD4 and Env in this example). Residues in CD4 that have been targeted by positive selection (red spheres) were mapped onto the co-crystal structures of the CD4 D1 domain in complex with both of its binding partners: MHC-II and Env (gp120) (Figure 4-4A-B). Although these CD4 residues appear to fall in a binding interface that is shared between both of these interaction partners, closer inspection reveals that positively selected residues track more closely with Env gp120 binding than with MHC-II binding (Figure 4-4C). Therefore, CD4 is predicted to be constant across species in its interaction with MHC-II, but variable between species in its interaction with HIV-1 and related viruses.

Furthermore, codons that have evolved under positive selection are often polymorphic within species (25, 42, 92, 122, 127, 190, 191). Many of the SNPs that were identified in *CD4* (gray boxes in Figure 4-4C) cluster around codons under positive selection, and there are 6 positions where SNPs exactly overlap with sites under selection (T15, N32, I34, N39, N52, and A55). This suggests that significant variation in CD4 function will exist between individuals of the same species.

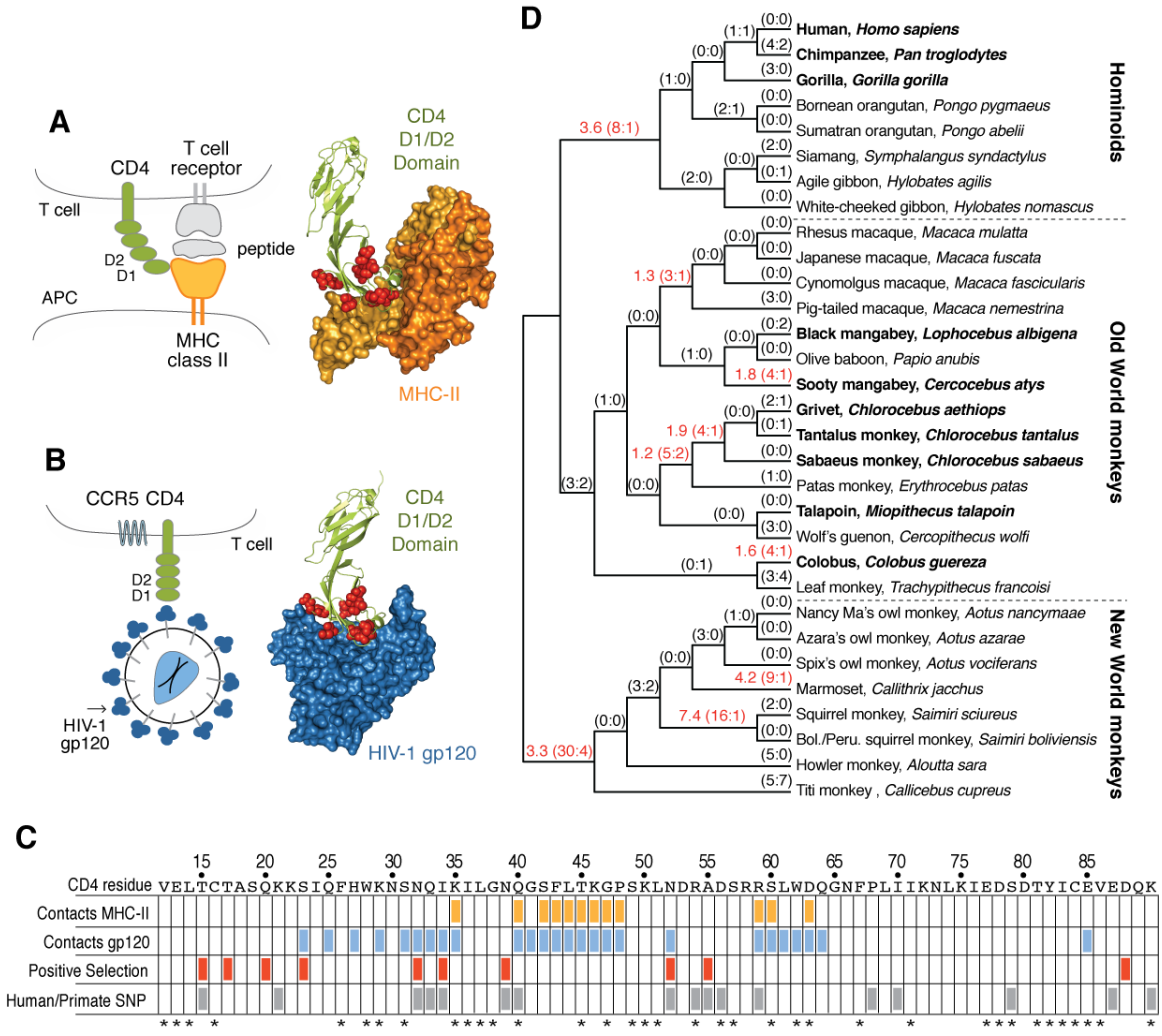


Figure 4-4: Evolution of the CD4 D1 domain.

Figure 4-4: Evolution of the CD4 D1 domain. (A) A schematic of the interaction between CD4 and MHC-II, alongside its co-crystal (PDB ID:1JL4) (192) where the sites under positive selection in CD4 (27) are represented by red spheres. APC = antigen presenting cell. **(B)** Schematic of the interaction between CD4 and HIV-1 gp120, alongside its co-crystal (PDB ID:1RZJ) (193) where the sites under positive selection in CD4 (27) are represented by red spheres. **(C)** A table showing amino acids in the CD4 D1 domains that contact gp120 (blue) or MHC-II (orange), are under positive selection (red; (27)), or are polymorphic for non-synonymous mutations in primate populations included in the present study (gray). Asterisks along the bottom denote residue positions that are completely conserved in the 31 primate species shown in panel D. Table format modified from (139). **(D)** Evolutionary analysis of the D1 domain of *CD4* showing the number of non-synonymous and synonymous mutations (in parentheses; N:S) predicted to have occurred along each branch of a 31 species primate phylogeny (194). Branches with red text have dN/dS >1, and this ratio is shown before the parentheses. This ratio cannot be calculated in cases where dS=0. Primate species shown in bold text are known to be naturally infected with a simian or human immunodeficiency virus (195).

But why does variation in CD4 function exist in New World monkeys, which have never been found to harbor lentiviruses? The newly generated *CD4* sequences, along with existing sequences, were used to reconstruct the ancestral *CD4* sequence that existed at each node of a 31 species primate phylogeny using PAML (100) (Figure 4-4D). On each branch of the tree, the ratios in parentheses represent the numbers of non-synonymous and synonymous substitutions (N:S) predicted to have occurred in the D1 domain. To detect selection, these values must be normalized to the number of opportunities that existed for each type of mutation in the DNA sequence of *CD4* (converting these raw counts to the rates dN and dS). The D1 domain experienced selection in favor of non-synonymous mutations (dN/dS > 1) along many branches of the tree (shown in red type). Some of the most extreme dN/dS values are found in the New World monkey clade. This observation, along with the presence of the potent TRIM5-CypA restriction factor in the *Aotus* genus of New World monkeys (196-198), suggests that this primate clade may have been antagonized by lentiviruses in their evolutionary past. The resulting host-virus arms race drove key residue positions in CD4 to become highly diverse both within and between species.

DISCUSSION

Here, we have identified the first nonhuman primate species that encodes a CD4 receptor compatible with Envs from primary, unadapted isolates of HIV-1. These permissive *CD4* alleles already exist in the genomes of living animals and, unlike transgenes made in the lab, can be manipulated through breeding techniques that do not

require transgenic approaches. A host-virus arms race appears to exist between *CD4* and lentiviral *env,* and this has driven the positive selection of *CD4* over evolutionary time leaving behind substantial sequence and functional variation in the *CD4* alleles encoded by different individuals of the same species. As such, we have demonstrated how individual monkeys can vary in their ability to support HIV-1 replication, and that gross conclusions made about the inability of certain species to support HIV-1 replication need to be revisited with population genetics in mind. This study serves as a powerful proof of principle that valuable nonhuman primate models may have been overlooked simply because individuals with the correct combination of alleles in their genetic background have not been identified.

Besides Spix's owl monkey, the other two owl monkey species that were examined, Nancy Ma's and Azara's owl monkeys, encode *CD4* alleles with D1 domains that are identical to that of permissive Spix's owl monkey *CD4* alleles (Figure 4-1), and therefore are likely to also support entry by circulating HIV-1 Envs. This suggests that there is still much to be discovered about the genetic capabilities of potential HIV-1 animal models.

Systematic exploration of genetic variation at the HIV-1 receptor and restriction factor loci in primate populations is now warranted. The literature suggests that there is rich genetic variation to be unearthed. SNPs at restriction factor loci have been shown to dramatically affect the outcome of infection experiments (28, 122, 160, 199-201). Less is known about polymorphism at host loci that encode the HIV-1 receptors, including *CD4* and *CCR5*. If future studies are successful at finding and combining permissive

restriction factor and receptor alleles into a single animal, additional genetic loci like the major histocompatibility complex (MHC) and Toll-like receptors will also have to be considered (202-205).

New World monkeys, like the Spix's owl monkey, are found in Central and South America and Mexico. They are generally small in size, have short gestation periods, and many are already being used as models for human diseases (206). New World monkeys became a subject of interest as potential models for HIV-1 pathogenesis when it was observed that their cellular blocks to infection are minimal. For instance, cells from various species of squirrel monkeys, marmosets, and tamarins are infected at high levels by VSV-G -pseudotyped HIV-1 vectors, suggesting that no restriction factor blocks exist after virus entry up to the stage of genome integration (207). Subsequent to this observation, primary cells from common marmosets (*Callithrix jacchus*) and squirrel monkeys (*Saimiri sciureus*) were found to support every phase of HIV-1 replication (reverse transcription, integration, transcription, translation, assembly, and budding) except for the entry of HIV-1 into cells (189).

The conclusion that cellular entry is the only barrier to HIV-1 replication in New World monkey cells has mostly held true. The one exception has been owl monkeys, which have been found to harbor three restriction factors active against HIV-1. The use

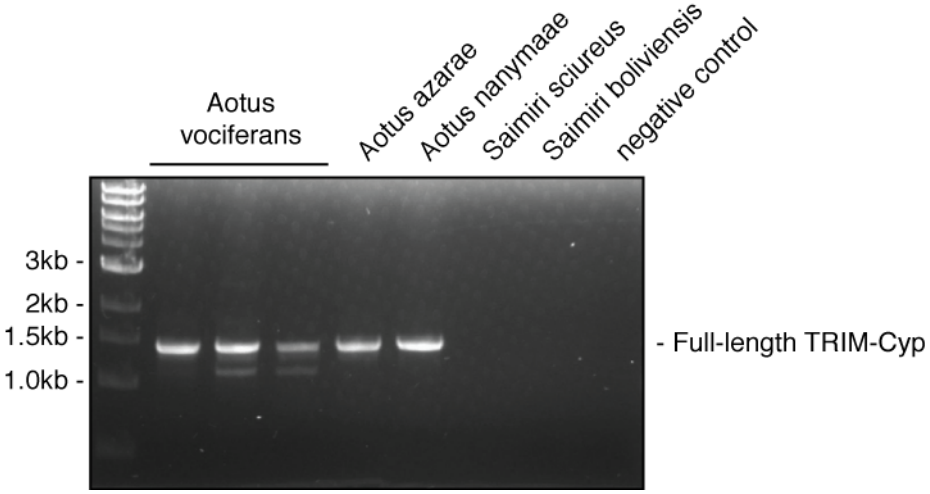


Figure 4-5: *Aotus vociferans* (Spix's owl monkey) express TRIM5-CypA. cDNA from three *Aotus vociferans* individuals along with other New World monkeys was used as a template in a PCR reaction to amplify full-length TRIM5-CypA. Primers were designed to amplify a 1,424 bp product that represents full-length TRIM5-CypA. Amplicons were confirmed via sequencing using PCR primers.

of owl monkeys, including Spix's owl monkey, as a model organism is not precluded by any of these, as it is possible to bypass these restrictions through modifications of HIV-1 that are independent of Env. First, all species of owl monkeys, including Spix's owl monkey (Figure 4-5), encode the TRIM5-CypA restriction factor (196, 197, 208). It is possible to bypass this restriction with a single amino acid mutation in HIV-1 capsid protein (G89V) (196, 197). The SIVmac capsid, which is used in all SHIV strains, also is not susceptible to owl monkey TRIM5-CypA (209). Second, some species of owl monkeys, including the Spix's owl monkey, have functional versions of the tetherin restriction factor that is not overcome by HIV-1 Vpu (176). Third, some species of owl monkeys express levels of APOBEC3G high enough to interfere with HIV-1 replication, but HIV-1 Vif can partially overcome this block (176).

SHIVs modified to replicate in Spix's owl monkey may allow experiments to be performed with viruses bearing clinically-relevant Envs derived from HIV-1. An even better scenario would be if functional receptor alleles could be identified in other New World monkey species that do not have these restriction factor blocks. In this case, an animal might even be identified that would harbor infection of unmodified HIV-1. Bioprospecting into the genomes of nonhuman primates housed in primate centers, particularly New World monkeys, could lead to the discovery of more desirable genotypes for HIV-1 research. Favorable alleles could then be combined through simple and tractable breeding techniques.

Chapter 5

The cyclophilin domain of RANBP2 exhibits species-specific interactions with lentiviral capsids

RANBP2/Nup358 is the largest component of the nuclear pore complex and the major constituent of the pore's cytoplasmic filaments. Human immunodeficiency virus type 1 (HIV-1) utilizes RANBP2 during the nuclear import of the pre-integration complex. Although the exact role of RANBP2 during this process is not fully understood, it is clear that the lentiviral capsid protein interacts with the cyclophilin domain of RANBP2 (RANCyp). We have previously documented an excess accumulation of amino-acid altering mutations and positive selection in RANCyp during the divergence of simian primate species. Here, we investigate how the genetic divergence in primate RANCyp affects interactions with diverse lentiviral capsids using a TRIM-fusion assay. The results show that the region in capsid that mediates interactions with a more well-characterized cyclophilin molecule, CypA, is also the same region that mediates interactions with RANCyp. Using chimeric capsid proteins we investigate how naturally occurring evolution in capsid between different primate lentiviruses has responded to divergence in RANCyp, and reveal that a recent cross-species transmissions of simian immunodeficiency virus (SIV) from chimpanzees into gorillas coincided with changes in capsid that allowed the exploitation of gorilla RANCyp. Ancestral reconstructions of RANCyp were then used to highlight important amino acid changes that have occurred

during the speciation of the great apes. These observations are in line with a scenario where *RANBP2* is a barrier to cross-species transmissions of primate lentiviruses in nature.

INTRODUCTION

Lentiviruses, like HIV-1, are unlike other families of retroviruses in that they are capable of infecting non-dividing cells. Although the lentiviral mechanism of infecting non-dividing cells is not fully understood, the major determinant is the capsid protein (142, 210). Replacing the capsid protein in a retrovirus that can only infect dividing cells with HIV-1 capsid confers the ability to infect non-dividing cells. However, there still remains some uncertainty as to how the capsid protein interacts with host proteins to allow access of the pre-integration complex into the nucleus. Recent whole-genome screens for host proteins required for HIV-1 infection revealed that several members of the nuclear pore complex, a massive multisubunit complex, which is composed of several copies of 30 nuceloporins embedded in the nuclear membrane and is responsible for selective nuclear import and export (211, 212), are exploited by lentiviruses to gain access to the nucleus (114-118, 213). Subsequently, various studies have confirmed the role of several nucleoporins in the nuclear import of the HIV-1 pre-integration complex (150, 214-216).

In particular, RANBP2 has generated interest as a key player in lentiviral nuclear import. RANBP2, also known as Nup358, is the major constituent of the nuclear pore's cytoplasmic filaments (217, 218). There are eight copies of RANBP2 at every nuclear

pore that mirror the octagonal symmetry of the pore itself (149, 219). It is a giant protein composed of 3225 amino acids and multiple protein domains, including a leucine-rich region, eight zinc-finger motifs, four Ran-binding domains, a C-terminal cyclophilin domain, and multiple FG-repeat domains that mediate cargo import and export through the nuclear pore (217, 218). RANBP2 was first implicated in HIV-1 replication when it was shown to regulate the shuttling of HIV-1 Rev protein between the cytoplasm and nucleus (220, 221). This protein also affects HIV-1 replication in two independent whole-genome siRNA screens (115, 116). Subsequently, a more detailed analysis confirmed a role for RANBP2 in the import of the HIV-1 pre-integration complex into the nucleus during the early phases of the viral life cycle (222). However, whether or not RANBP2 is an essential host factor for HIV-1 remains debatable. This is due in part to the redundancy in nuclear import pathways used for HIV-1 replication (223). One observation that has resolved this issue is that although HIV-1 can use redundant pathways for import, these alternative pathways lead to suboptimal integration sites for the provirus (223). Therefore, even if HIV-1 uses an alternative pathway to enter the nucleus, this is not a sufficient condition for productive infection. Another confounding factor is that not all lentiviruses are affected by RANBP2 knockdown (101, 224), which brings into question the significance of RANBP2 in the nuclear import of lentiviral pre-integration complexes.

We have documented signatures of positive selection in primate *RANBP2* (27). These signatures have previously been shown to identify regions of host proteins that are critical for modulating interactions with primate lentiviruses (11, 126-128). The presence

of this signature in *RANBP2* is in support of its critical role in the nuclear import of lentiviruses. More specifically, we identified the cyclophilin domain of RANBP2 (RANCyp) to contain several amino acid residues with a heightened rate of evolution. In fact, it was recently shown that RANCyp directly interacts with HIV-1 capsid to mediate the nuclear entry of the pre-integration complex (101, 150), and the co-crystal structure of this interaction reveals that the amino acids in RANCyp that were identified to be evolving under positive selection are either at or near the interaction region (225). In this study, we aim to understand how genetic divergence in primate RANCyp affects interactions with lentiviral capsids.

The RANCyp domain of RANBP2 is homologous to another very important protein in HIV-1 biology, cyclophlin A (CypA) (226-231). HIV-1 capsid, along with some other lentiviral capsids, interact with CypA to stabilize the capsid core during cellular entry (232, 233). Comparing the co-crystals of CypA-capsid and RANCyp-capsid shows a high level of similarity, which suggests a conservation of function (225, 234). However, why stability of the capsid core would be necessary at the nuclear pore in unclear, because it is generally accepted that the core must partially disassemble for the pre-integration complex to gain access to the nucleus. One explanation is that although the prolyl isomerase activity of CypA mediates stability of the capsid core, the same prolyl isomerase activity of RANCyp may promote the disassembly of the capsid core at the nuclear pore complex (214). Adding to the complexity of cyclophilin interactions that lentiviruses must juggle is the presence of TRIM-CypA fusion molecules in some primate species (196, 235, 236). These molecules are antiviral in nature and have resulted from

the natural fusion of *TRIM5* and *CypA* genes. Whereas CypA and RANCyp promote the replication of lentiviruses, TRIM-CypA is a potent inhibitor of lentiviral replication and therefore drastically affects how lentiviruses can maneuver the mutational landscape when adapting to new species (28, 237, 238). Because all of these interactions depend on capsid and occur early in the viral life cycle, we became interested in further characterizing the role of RANCyp during nuclear import and how primate lentiviruses maintain interactions with RANCyp while preserving the interaction state with other cyclophilin domains.

In the current study, how various primate RANCyp alleles interact with diverse lentiviral capsids was investigated. We began by comparing and contrasting the evolutionary modes of *CypA* and *RANCyp* to gain insight into how these two molecules have been shaped during the divergence of primates. Using a TRIM-fusion assay in combination with capsid mutants, we functionally characterized the genetic divergence of primate RANCyps, identified the region in capsid that is critical for interaction with RANCyp, revealed amino acids in RANCyp that mediate interactions with capsid, and showed that a natural cross-species transmission of a primate lentivirus coincided with adaptation to utilize the RANCyp in the new host species. Together, these results lend credence to the role of RANBP2 during the HIV-1 life cycle and are consistent with a scenario where some cross-species transmissions of primate lentiviruses may require the adaptation of capsid to utilize the *RANBP2* allele of its new host.

MATERIALS AND METHODS

Primate biomaterials used

Primary and immortalized primate cell lines from primate species were grown in standard media supplemented with 15% fetal bovine serum at 37ºC and in 5% CO_2.

Primate *RANBP2* and *CypA* gene sequences

Human Refseq sequences for *RANBP2 and CypA* were obtained from the NCBI nucleotide database. Chimpanzee, orangutan, rhesus macaque, and marmoset gene sequences were obtained from the UCSC genome database (http://genome.ucsc.edu/) using the BLAT alignment tool. Genes were sequenced from 22 additional primate species, and from chimpanzee, orangutan, rhesus and marmoset in instances where the genome-project sequences were of poor quality. PCR or RT-PCR was performed from total RNA, gDNA, or cDNA with SuperScript III One-Step RT-PCR system with Platinum Taq (Invitrogen, #12574-018), PCR SuperMix High Fidelity (Invitrogen, #10790-020), or Phusion High Fidelity PCR Master Mix (NEB, #F-531S). Primers were designed in conserved regions of the 5' or 3' UTR in order to amplify full open reading frames.

PAML analysis

Codon models were tested with codeml in the PAML 4.1 software package (93). To detect selection, multiple alignments were fit to the NSsites models M8a (neutral model, codon values of dN/dS fit to a beta distribution plus an extra codon class fixed at

dN/dS = 1) and M8 (positive selection model, similar to M8a but with the extra class allowed to be dN/dS >1). A likelihood ratio test was performed to assess whether permitting codons to evolve under positive selection gives a significantly better fit to the data (model comparison M8a vs. M8). In situations where the null model could be rejected ($p < 0.05$), posterior probabilities were assigned to individual codons belonging to the class of codons with dN/dS > 1. Residues under positive selection were mapped onto existing crystal structures using MacPyMol (v.0.99; http://pymol.sourceforge.net/). A branch model was used to calculate the most likely locations of non-synonymous and synonymous changes (N:S in Figure 5-1) along each lineage in the primate phylogeny used in this study.

TRIM-CypA and TRIM-RANCyp expression constructs

HA-tagged owl monkey TRIM-CypA in the pLPCX expression vector was a kind gift from Michael Emerman. TRIM-RANCyps were constructed by generating fragments with 20-25bp overlapping regions of owl monkey TRIM-CypA and RANCyp. Overlapping fragments were spliced together using a PCR reaction and each fragment as a template with outside flanking primers. Constructs were TA-cloned into pCR4 (Invitrogen, #K4575-01). An N-terminal HA tag was added using a PCR reaction and these tagged constructs were TA-cloned into the Gateway entry plasmid pCR8 (Invitrogen, #K2500-20). An LR Clonase II reaction (Invitrogen, #11791-100) was used to move these constructs into a Gateway-converted pLPCX retroviral vector (Clontech, # 631511). Ancestral TRIM-RANCyp alleles were generated using PfuTurbo DNA

polymerase (Stratagene, #600250). Parental pLPCX plasmids were used as a template along with primers containing the mutations of interest.

Generation of stable cell lines

To generate cell lines that stably express owl monkey TRIM-CypA and TRIM-RANCyps, retroviral vectors were used to transduce CRFK cells (ATCC). To generate the retroviral vectors, 293T cells were seeded at a concentration of $1x10^6$ cells/well in a 6-well dish. 24 hours later each well was transfected with 2 mg pLPCX construct (empty or encoding the gene fragment of interest), 1 μg pCS2-mGP encoding MLV gag-pol (142), and 0.2 μg pC-VSV-G at a final 1:3 ratio of DNA to TransIT-293 (μg DNA : ml TransIT-293). Supernatants were collected after 48 hours, passed through a 0.2 μm filter, and used to infect CRFK cells grown in DMEM supplemented with 10% fetal bovine serum, 100 IU/ml penicillin, 100 μg/ml streptomycin, and 2 mM L-glutamine. After 24 hours, media containing 8 μg/ml puromycin was added to select for transduced cells. Expression of TRIM-CypA and TRIM-RANCyp constructs was detected by Western blot.

Antibodies and Western blot analysis

Cell lines were grown to confluency in a 6-well dish, collected using a cell scraper, and lysed in a buffer containing 150 mM NaCl, 50 mM Tris-HCl (pH 7.4), 1% NP-40, and Complete protease inhibitor (Roche, #11836170001). After quantitation of protein concentration using a Bradford assay, 30 μg of whole cell extract was resolved

using a 10% polyacrylamide gel and transferred to a nitrocellulose membrane. HA-tagged constructs were detected using a 1:5000 dilution of anti-HA-peroxidase antibody (Roche, #12013819001). Maturation of Gag protein was monitored using a 1:1000 dilution of anti-p24 (AIDS reagent database, #183-H12-5C). β-actin was also detected for a loading control using a 1:1000 dilution of mouse anti-β-actin (Santa Cruz, #sc-47778). A 1:10,000 dilution of donkey mouse-specific horseradish peroxidase-conjugated antibody (Thermo Scientific, #32430) was used as a secondary probe. Blots were developed using the ECL Plus detection reagent (GE Healthcare, #RPN2132).

Retroviral integration assays

Viruses for single-cycle infection assays were packaged in 293T cells by co-transfection of plasmids encoding viral proteins and VSV-G, along with a transfer vector, as follows: HIV-1 and cyclophilin binding loop mutants (pMDLg/pRRE, pRSV-Rev, pMD2.G, pRRLSIN.cPPT.PGK-GFP.WPRE; all available from Addgene), FIV (pFP93 (144), pC-VSV-G, pGIN-SIN:GFP (144)). After 48 hours, supernatant containing viruses was harvested, filtered, and frozen. For infection assays, CRFK stable cells lines were plated at a concentration of $7.5x10^4$ cells/well in a 24-well plate and infected with HIV-1, cyclophilin binding loop mutants, or FIV the following day. Two days post-infection, cells were fixed, washed, resuspended in PBS supplemented with 1% FBS, and analyzed by flow cytometry for expression of GFP using the BD Bioscience Fortessa cell analyzer. Experiments were performed out in triplicate, and results were confirmed with independent experiments.

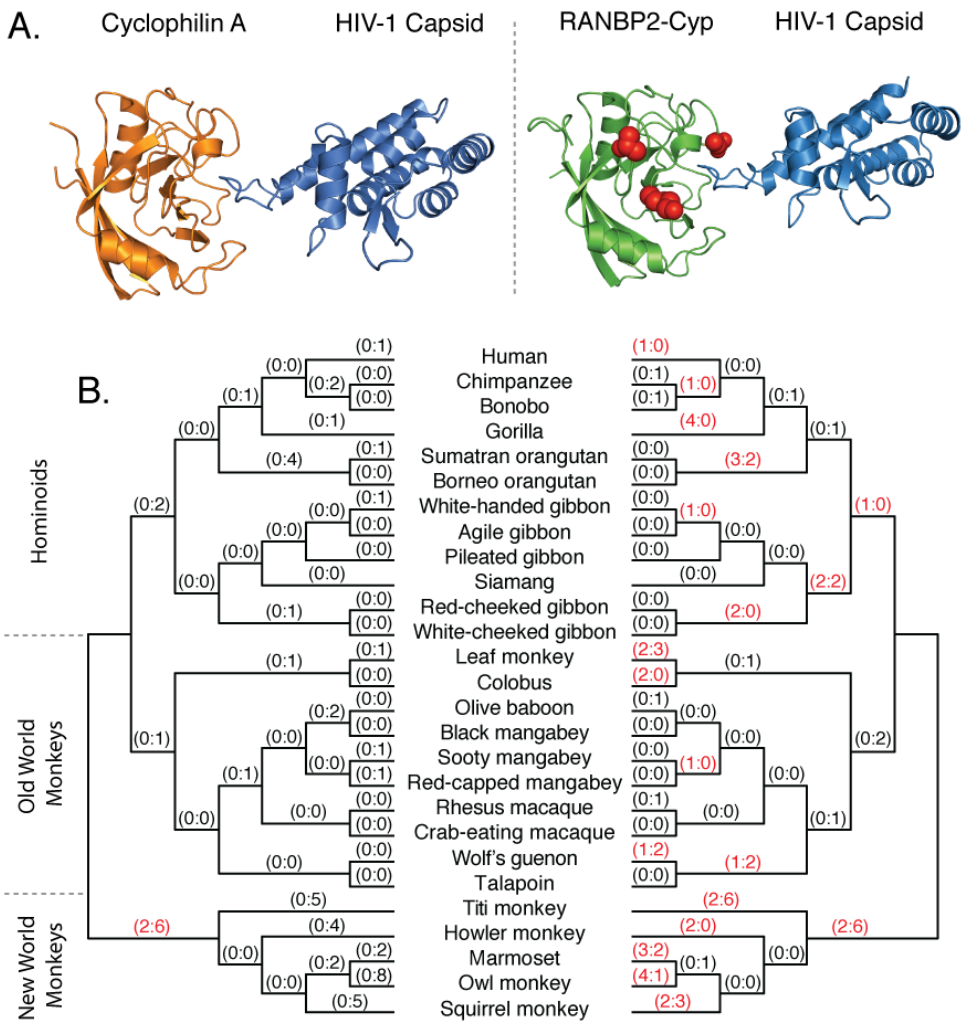
Cyclophilin-binding loop capsid mutants

pMDLg/pRRE expressing HIV-1 *gag-pol* was used as a template for site-directed mutagenesis using PfuTurbo DNA polymerase (Stratagene, #600250).

RESULTS

CypA and RANCyp have contrasting evolutionary histories in primates

CypA and RANCyp interact with HIV-1 capsid in a highly similar fashion (Figure 5-1A) (225, 234). Alignment of homologous regions in CypA-capsid and RANCyp-capsid co-crystals results in an RMSD = 1.323Å, indicative of a conserved interaction. This observation is consistent with the functional conservation of HIV-1 capsid prolyl isomerase activity of both CypA and RANCyp, however, it conflicts with the earlier report that *RANBP2* has evolved under positive selection in the primate lineage (27). To further investigate this paradox, we extended the analysis of the RANCyp domain and performed a parallel evolutionary analysis of CypA. To do this, we generated 7 additional primate RANCyp sequences and 22 primate CypA sequences. This allowed us to construct alignments of CypA and RANCyp, each composed of a total of 27 primate species (Figure 5-1B).



C.

Test for positive selection using PAML models M8 vs M8a

Data set	-2ΔlnL	*p*-value	% sites with dN/dS>1	Codons P > 0.90
CypA	0.0	$p = 1.0$	na	na
RANCyp	8.1	$p < 0.005$	9.3	T93, Q103, K117

Figure 5-1: Contrasting evolutionary modes of CypA and RANCyp.

Figure 5-1: Contrasting evolutionary modes of CypA and RANCyp. (A) Co-crystal structures of CypA bound to HIV-1 capsid (left, pdb:1AK4) and RANCyp bound to HIV-1 capsid (right, pdb:4LQW). Amino acid residues under positive selection are shown as red spheres. **(B)** Phylogenetic tree of the primates used in the evolutionary analysis. Numbers shown in parenthesis are (number of non-synonymous changes : number of synonymous changes) as calculated using an ancestral reconstruction in the codeml program of PAML. Lineages with numbers if red have experienced non-synonymous changes. **(C)** Results of a site-based evolutionary analysis using the codeml program of PAML. Here, a model that allows positive selection (M8) was compared against a null model that does not allow positive selection (M8a). Amino acid numbering is relative to homologous positions in CypA.

The evolutionary history of *CypA* and *RANCyp* was studied by employing maximum likelihood codon models of evolution as provided in the PAML package (239). First, we constructed ancestral sequences at each node of the primate phylogeny to track where non-synonymous and synonymous DNA mutations have occurred (Figure 5-1B) (194). *CypA* is almost entirely conserved in the primate lineage, with only two non-synonymous mutations (E84D, A117V) occurring on the lineage leading to New World monkeys. Neither of these mutations have been implicated in HIV-1 biology, although most research in this context has been performed on the TRIM5-CypA fusion molecule (240-243). On the other hand, *RANCyp* has a dynamic evolutionary history. Non-synonymous mutations have accumulated on 19 branches throughout the primate phylogeny (branches with red labeling in Figure 5-1B), often in excess of synonymous mutations. Nearly half of these branches are located in the hominoid clade, where the most rapidly diverging branch leading to gorilla is located. Site-based tests were also used in PAML to identify individual codon positions that are evolving under positive selection (see Methods). There was no evidence of positive selection in *CypA*, however, an intense signal of positive selection was identified in *RANCyp* that was consistent with a previous analysis of this domain (Figure 5-1C) (27). Therefore, *Cyp*A and *RANCyp* have contrasting evolutionary histories even though their structures and interactions with HIV-1 capsid are nearly identical.

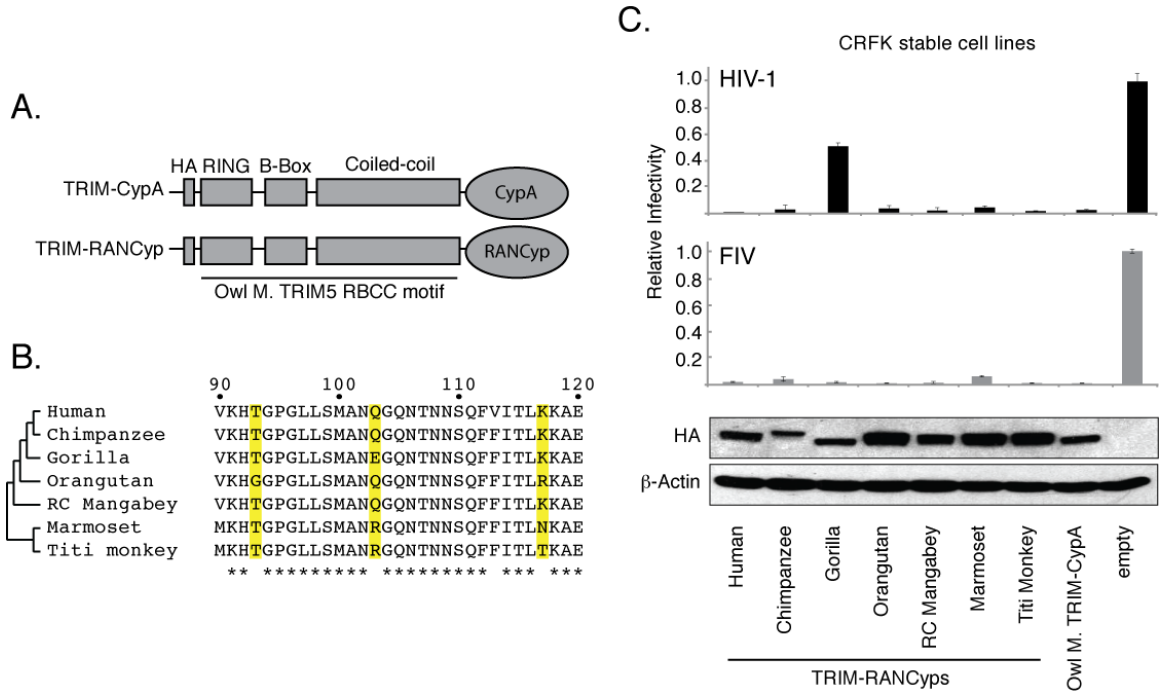


Figure 5-2: TRIM-RANCyp fusion proteins differentially restrict divergent lentiviruses. (A) Protein schematic of TRIM-fusion constructs that are stably expressed in CRFK cells. **(B)** Protein alignment of the RANCyp alleles that were constructed were placed in a phylogenetic context. The region shown encompasses the three amino acid positions under positive selection, which are each highlighted in yellow. Amino acid numbering is relative to homologous positions in CypA. **(C)** Single-cycle infection assay of CRFK stable cell lines using either HIV-1 or FIV pseudotyped virus containing a GFP reporter. Represented as relative infectivity relative to the empty vector control. Expression of each fusion construct was observed using Western blotting of whole cell lysates.

Divergence in primate RANCyp affects binding to lentiviruses

The evolutionary history of primate *RANCyp* suggests that there may be functional divergence in this domain with respect to capsid binding. To test this, a RANCyp-capsid interaction assay that has been utilized to study host interactions with capsid was used (Figure 5-2) (101, 158, 244, 245). A chimeric protein that consists of the human RANCyp domain fused with the core 'RBCC' motif of TRIM5α (human TRIM-

RANCyp) was shown to restrict HIV-1 infection similarly to the owl monkey TRIM-CypA allele, indicative of a human RANCyp-capsid interaction (101). We recapitulated these results and also generated six additional HA-tagged TRIM-RANCyp alleles representing various primate RANCyps (Figure 5-2A,B). These particular primate species were chosen to encompass the extent of sequence divergence in the region containing amino acid residues under positive selection. Stable cell lines expressing these TRIM-RANCyp alleles were generated in CRFK cells using a lentiviral vector system.

These TRIM-RANCyp cell lines were then challeneged with single-cycle, VSV-G pseudotyped, HIV-1 (Figure 5-2C). As expected, human TRIM-RANCyp and owl monkey TRIM-CypA interact with and restrict the HIV-1 capsid. The other 6 TRIM-RANCyp alleles behaved similarly to human TRIM-RANCyp, with the exception of the gorilla allele. This allele only restricted HIV-1 about 2-fold, which suggests that gorilla RANCyp does not interact with HIV-1 capsid as well as the other primate RANCyps tested. To be certain that gorilla TRIM-RANCyp is actually an active molecule and the lack of effect on HIV-1 is not simply due to misfolding of the protein, the panel of primate TRIM-RANCyp alleles was tested against a feline immunodeficiency virus (FIV) construct and confirmed to interact with and restrict FIV capsid (Figure 5-2C). Therefore, gorilla TRIM-RANCyp is a fully functional molecule that does not restrict HIV-1, but does restrict FIV capsid. Because there are only a total of 5 amino acid differences between human and gorilla TRIM-RANCyp, it is clear that the significant difference between these alleles in the infection assays is due to a small number of mutations in the

RANCyp domain. These data suggest that HIV-1 would be incapable of utilizing the nuclear pore of a gorilla cell via exploitation of the gorilla RANBP2 cyclophilin domain.

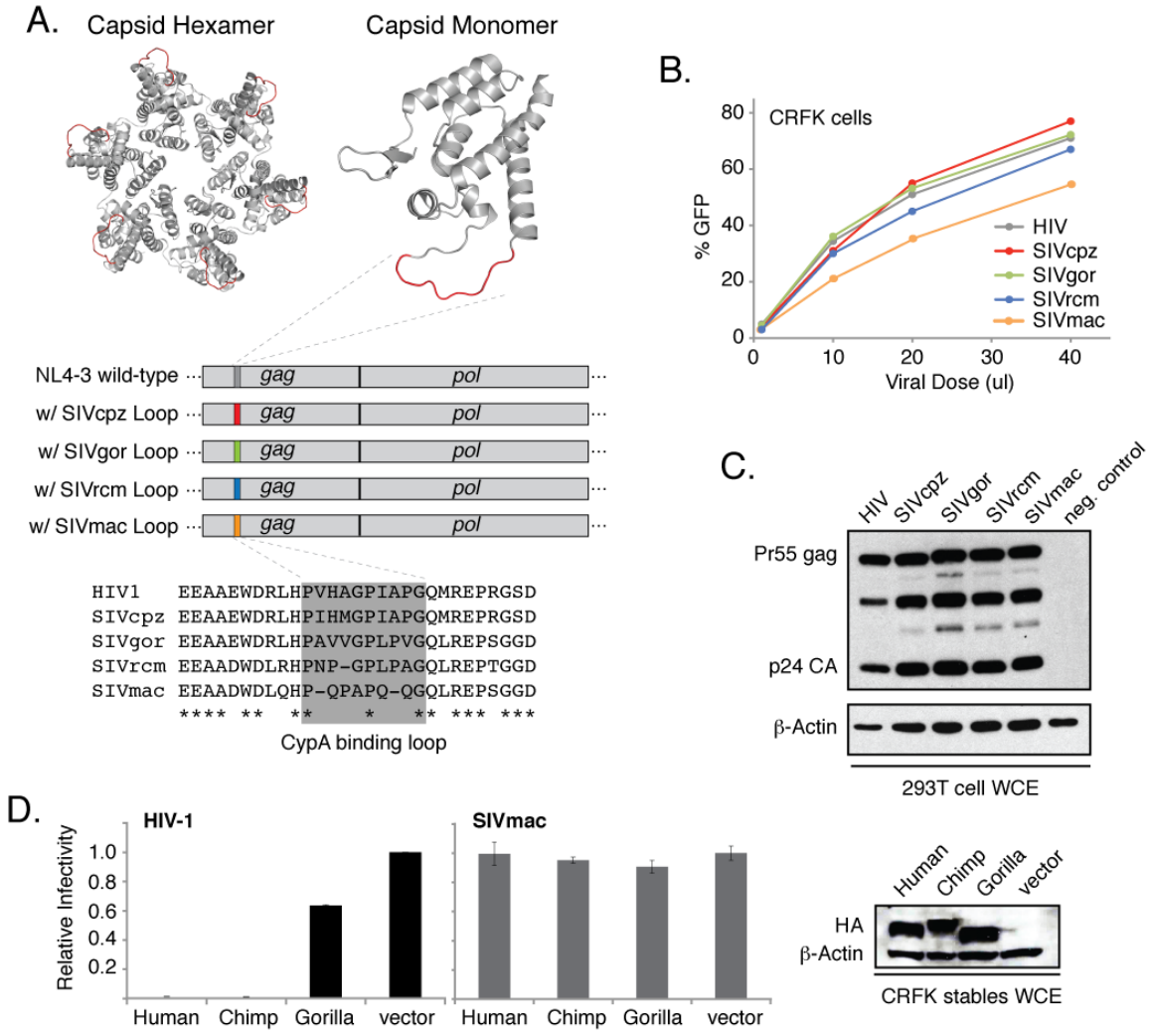


Figure 5-3: Cyclophilin binding loop of capsid mediates interaction with RANCyp.

Figure 5-3: Cyclophilin binding loop of capsid mediates interaction with RANCyp. **(A)** Crystal structures of hexameric (pdb:3GV2) and monomeric (pdb:1AK4) HIV-1 capsid forms. Region shown in red is the cyclophilin A binding loop (top). Region of *gag-pol* genomic region where the cyclophilin A binding loop is located (middle). An alignment of various lentiviral capsids used to generate mutants. 'CypA binding loop' is highlighted in gray (bottom). **(B)** Titering of mutant viruses on CRFK reporter cells. **(C)** Western blot of capsid mutants using an anti-p24 antibody to show proper maturation of mutant capsids. *Gag-pol* plasmids were transfected into 293T cells, and whole cell extract (WCE) was harvested at 48 hours post transfection. **(D)** Single-cycle infection assay using indicated TRIM-RANCyp stable cells lines and viruses that have the indicated 'CypA binding loop' in the HIV backbone. Representated as relative infectivity relative to empty vector control. A Western blot is also shown for the expression of the TRIM-RANCyp constructs in the stable cell lines.

CypA binding loop in capsid affects RANCyp interactions

After establishing that divergence in primate *RANCyp* can affect binding to capsid, we established the genetic determinants of this interaction in capsid sequences. At this point, we focused exclusively on the human, chimpanzee, and gorilla RANBP2 alleles for two reasons. First, a phenotypic difference with the gorilla allele was observed in the infection assays. Second, all three of these species have extant lentiviruses circulating in their populations and thus presents a convenient system for understanding the potential role of RANBP2 during cross-species transmissions (195). In relation to this point, chimpanzees harbor SIVcpz, which is currently thought to have been the progenitor of HIV-1 in humans and SIVgor in gorillas (246). With these considerations in mind, how genetic differences in lentiviral capsids affect binding to RANCyp was investigated.

We hypothesized that the CypA binding loop in capsid might also be the genetic determinant for RANCyp binding. To test this, mutant capsid chimeras were generated by replacing this 10 amino acid CypA binding loop in HIV-1 with the homologous region from other lentiviruses (Figure 5-3A, CypA binding loop is red in crystal structures). These mutants were made via site-directed mutagenesis of the HIV-1 *gag-pol* expression vector from the pseudotyping system. To ensure the viability of these mutants, single-cycle HIV-1 virus harboring the mutant CypA binding loops was produced and titered on CRFK cells (Figure 5-3B). Proper Gag maturation was confirmed by probing 293T producer cell lysates with a p24-specific antibody (Figure 5-3C). No significant

difference in titers or Gag maturation were observed, therefore these mutant viruses were used for downstream analyses.

It was previously shown that SIVmac does not utilize RANBP2 for entry into the nucleus (101). Therefore, if the CypA binding loop in capsid mediates interactions with RANBP2, then the HIV/SIVmac chimeric capsid should not bind RANCyp. Indeed, when the human, chimpanzee, and gorilla TRIM-RANCyp harboring cell lines were challenged with HIV/SIVmac chimeric virus, the virus was uninhibited (Figure 5-3D). This indicates that the 10 amino acid patch in capsid known as the CypA binding loop completely determines interactions with TRIM-RANCyp, and also suggests that this region of capsid could be critical for interactions with RANCyp in the context of full-length RANBP2 at the nuclear pore. Further, the assay used here has been shown to recapitulate both positive and negative interactions between RANCyp and lentiviral capsids.

Natural evolution of capsid correlates to RANCyp utilization

We now turn to understand how the dynamic interaction of RANCyp and lentiviral capsids has played out in a natural context by studying how this interaction has evolved during known cross-species transmissions. There are currently over forty primate species that are naturally infected with extant lentiviruses (247, 248). Collectively, these

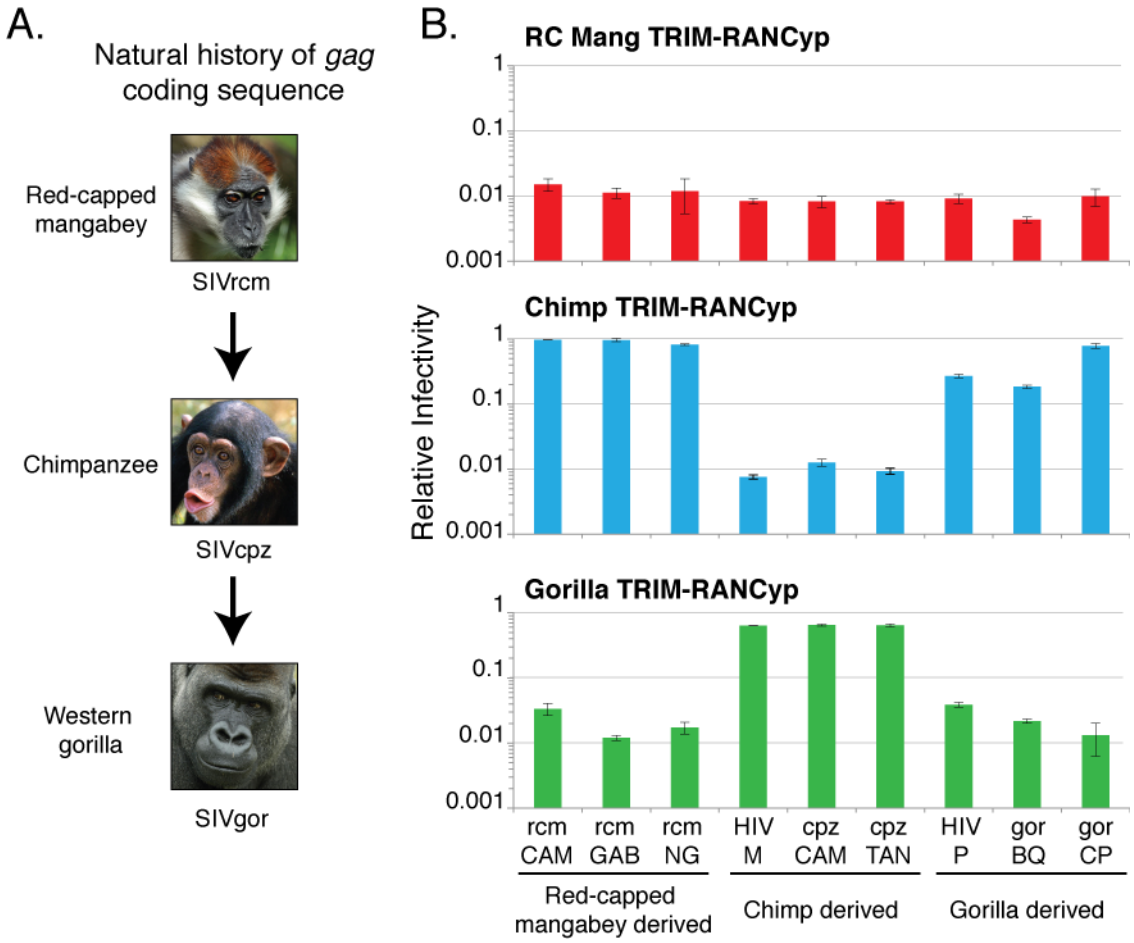


Figure 5-4: Capsid interaction with RANCyp is a potential barrier to cross-species transmission of lentiviruses. (A) Schematic of the natural cross-species transmissions of lentiviruses that have occurred amongst the species shown. Only the *gag* coding sequence has been taken into account. **(B)** Single-cycle infection assays of three primate TRIM-RANCyp cells lines using mutant capsids with the indicated 'CypA binding loop' inserted into the HIV-1 backbone. Represented as relative infectivity relative to empty vector (empty vector data not shown).

SIV strains constitute a large reservoir for cross-species transmissions amongst their primate hosts in Africa (249-251). Here, we will focus on two well-documented transmissions that have preceded the emergence of HIV-1 (Figure 5-4A). First, SIVcpz in chimpanzees emerged as a recombination event of SIV strains from red-capped

mangabeys (SIVrcm), greater spot-nosed monkeys (SIVgsn), mustached monkeys (SIVmus), and mona monkeys (SIVmon) (252). However, the entire region of the SIVcpz *gag* coding sequence descended from SIVrcm, and so for the purpose of studying the RANCyp-capsid interaction, SIVrcm can be considered as the most recent ancestor of SIVcpz. Second, SIVgor in gorillas emerged after SIVcpz infected and adapted to gorilla populations (246, 253). These two events are diagrammed in Figure 5-4A and will serve as the prime motivation for the following experiments.

Using the TRIM-RANCyp assay described above we studied how the red-capped mangabey, chimpanzee, and gorilla alleles interact with capsid mutants (HIV-1 backbone with chimeric CypA binding loops) representing SIVrcm, SIVcpz, SIVgor, and HIV-1. More specifically, the TRIM-RANCyp alleles were challenged with three red-capped mangabey derived strains (SIVrcmCAM, SIVrcmGAB, SIVrcmNG), three chimpanzee-derived strains (HIV group M, SIVcpzCAM, SIVcpzTAN), and three gorilla-derived strains (HIV group P, SIVgorBQ, SIVgorCP) (Figure 5-4B). All capsid variants are recognized and inhibited by the red-capped mangabey TRIM-RANCyp allele. This shows that sequence divergence in capsid does not affect its ability to bind red-capped mangabey RANCyp. However, chimp and gorilla TRIM-RANCyps each have distinct interaction phenotypes with the various capsid mutants. For example, chimpanzee TRIM-RANCyp interacts only with chimp-derived strains and gorilla TRIM-RANCyp interacts with only red-capped mangabey and gorilla derived strains. Two points of interpretation can be made. First, all three of these TRIM-RANCyp alleles interact with capsids derived from that same host (i.e. red-capped mangabey TRIM-RANCyp interacts with red-capped

mangabey derived SIV strains, chimpanzee TRIM-RANCyp interacts with chimpanzee derived SIV strains). This is to be expected if SIVrcm, SIVcpz, and SIVgor utilize RANBP2 during their life cycles to gain access to the nucleus. Another important point is evident only when considering the transmission of SIVcpz from chimpanzees into gorillas, which eventually evolved to become SIVgor. In this case, the SIVcpz strains were adapted to interact with chimpanzee RANCyp, but not gorilla RANCyp. However, after evolving to become SIVgor, the virus gained the ability to interact with gorilla RANCyp and lost the ability to interact with chimpanzee RANCyp. Our interpretation is that during the course of cross-species transmission, the ability to utilize RANBP2 may have been critical enough to warrant adaptation to gain this function. The net result was adaptation to utilize the RANBP2 of a new host at the expense of losing this function in the previous host.

Next we tested how the zoonotic events that lead to the emergence of HIV-1 might have been affected by divergence in *RANCyp*. The main group of HIV-1 emerged into the human population early in the 20th century (254), and since then several independent zoonoses have occurred that have led to the emergence of various HIV-1 strains (195). All of the known strains of HIV-1 were passed from either chimpanzees or gorillas into humans. Therefore, we proceeded to test how the RANCyp-capsid interaction was shaped during these events. We found that capsid mutants representing SIVcpz and SIVgor sequences were already capable of binding human RANCyp (Figure

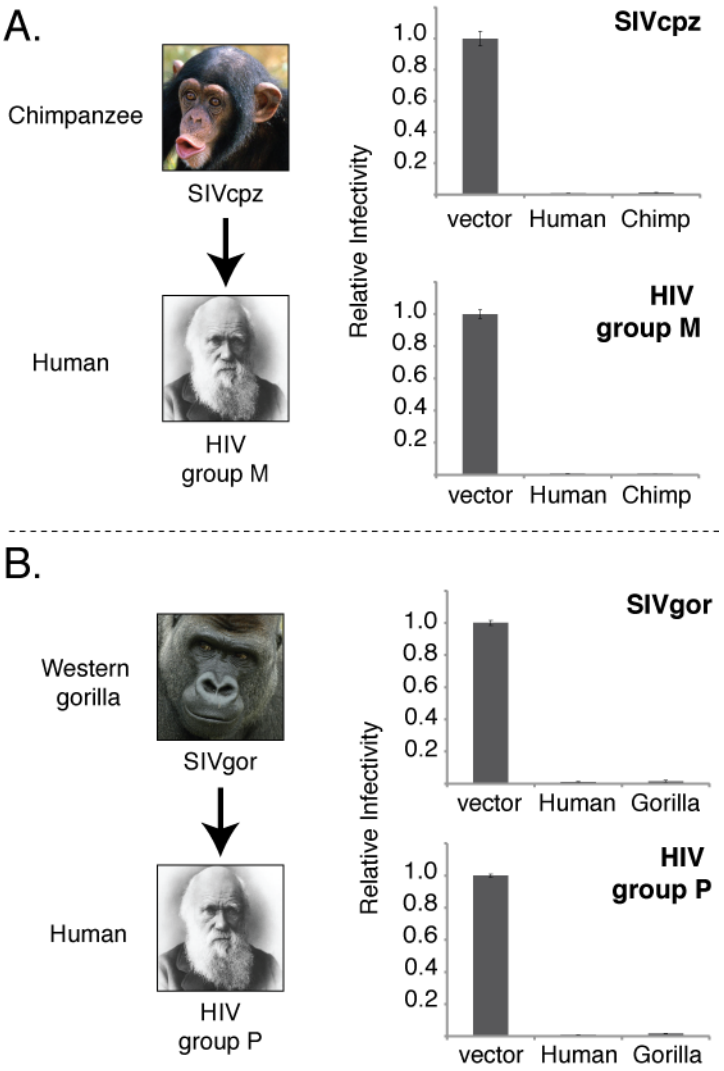


Figure 5-5: Capsid interaction with RANCyp was conserved during SIV zoonosis. **(A)** Schematic of the zoonotic event that led to the emergence of HIV-1 group M (left). Single-cycle infection assay of TRIM-RANCyp cells lines using mutant capsids with the indicated 'CypA binding loop' inserted into the HIV-1 backbone. Represented as relative infectivity relative to empty vector (right). **(B)** Schematic of the zoonotic event that led to the emergence of HIV-1 group P (left). Single-cycle infection assay of TRIM-RANCyp cells lines using mutant capsids with the indicated 'CypA binding loop' inserted into the HIV-1 backbone. Represented as relative infectivity relative to empty vector (right).

5-5). Additionally, this interaction was maintained as SIVcpz and SIVgor evolved into HIV group M and HIV group P, respectively. Therefore, adaptation to utilize human

RANBP2 was not an obstacle met by SIVcpz and SIVgor strains that crossed the species barrier into humans.

Ancestral reconstruction of great ape *RANCyp* identifies critical residues

The human, chimpanzee, and gorilla *RANCyp* alleles are all very similar, with at most five amino acid differences given any pairwise comparison. These differences are shown in both a phylogenetic context and an amino acid alignment (Figure 5-6A,B). The ancestor of this clade has been reconstructed and has also been included it in the Figure 5-6B alignment. Each species has accumulated its own unique mutations since divergence from the ancestor and are shown in parentheses next to the phylogenetic tree. Residues V113, H121, and Q103 have recently been shown to make important contacts in the RANCyp-capsid co-crystal (225), and therefore divergence at these sites amongst the primate alleles is likely to affect capsid interactions.

To test this we generated ancestral TRIM-RANCyp alleles and stably expressed them in CRFK cells (Figure 5-6C). Three gorilla-specific mutations (G75R, K82R, and Q103E) were used because the effects of the other mutations (F113V, H121L, and K149E) can be inferred from previous experiments, as the ancestral *RANCyp* allele is identical to the red-capped mangabey *RANCyp* allele (Figure 5-2 and 5-4). All possible single, double, and triple mutants using the gorilla mutations were generated relative to the ancestral allele and tested their interactions with various capsid mutants using the TRIM-fusion assay (Figure 5-6D). All representative capsid mutants that were tested (HIV-1, SIVcpz, SIVgor, and FIV) bound the ancestral *RANCyp* allele. Interestingly,

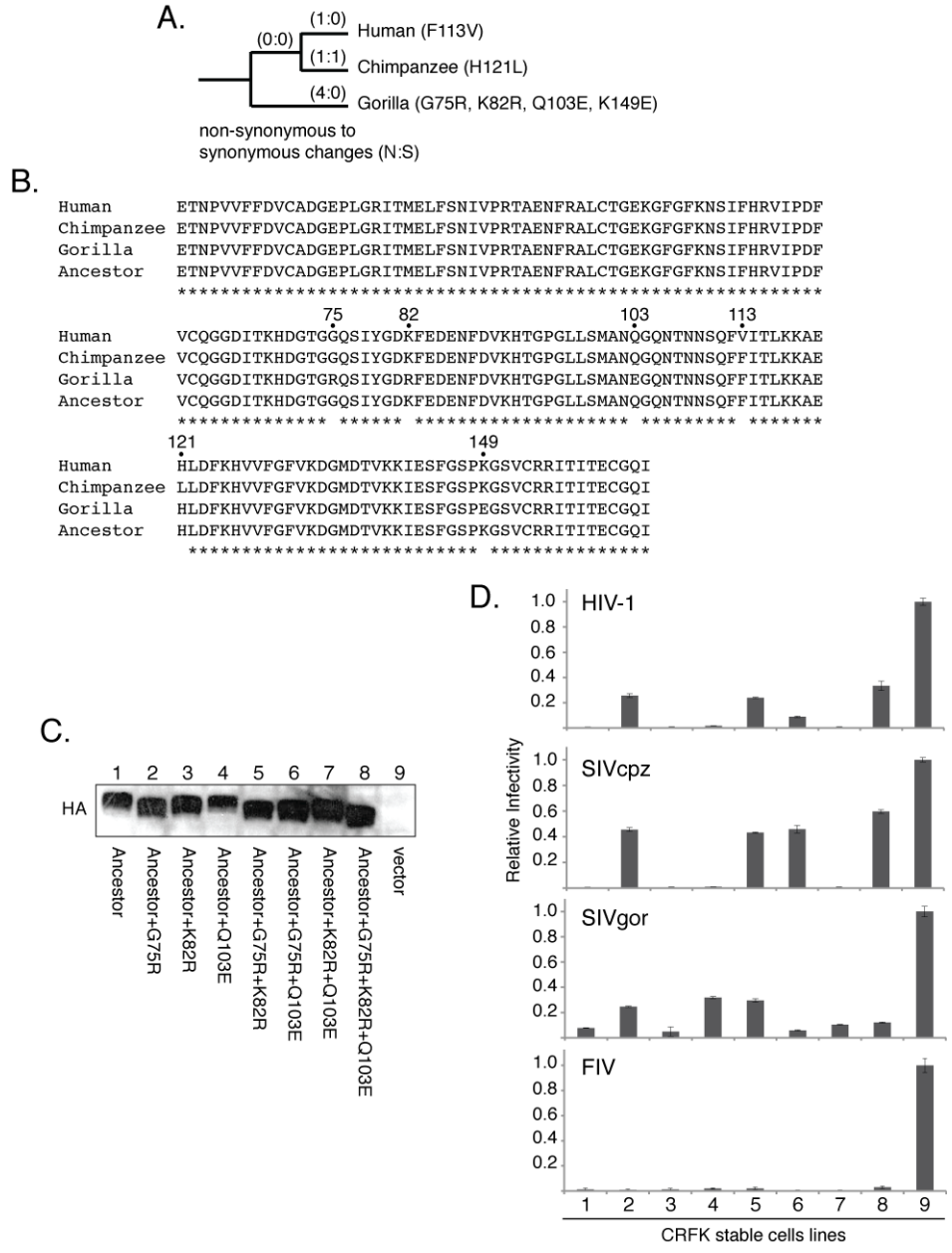


Figure 5-6: Ancestral reconstruction of RANCyp reveals critical amino acid residues. (A) Amino acid changes along the lineages leading to human, chimpanzee, and gorilla RANCyp placed in a phylogenetic context. **(B)** Alignment of RANCyp alleles used to construct the ancestral allele. Amino acid numbering is relative to CypA. **(C)** Western blot showing the expression and labels for the ancestral RANCyp alleles that are stably expressed in CRFK cells. **(D)** Infection of the ancestral RANCyp cell lines using mutant capsids with the indicated 'CypA binding loop' inserted into the HIV-1 backbone.

none of the mutations tested affected binding to FIV capsid, which was invariably maintained. Combined with data from Figure 5-1, all the divergence in primate RANCyp discussed in this study has no effect on binding FIV capsid. Therefore, the mutations that have accumulated in primate *RANCyp*s seem to be in response to primate lentiviral capsids. All three of the gorilla-specific mutations had significant effects, although sometimes in different contexts. For example, the G75R mutation always results in a loss of binding to HIV-1 and SIV, whereas the K82R had an effect on SIVgor capsid binding in conjunction with the Q103E mutation. The SIVgor capsid mutant is the only variant that is able to bind the triple mutant, which is expected since this virus has adapted to use gorilla RANCyp. This is also the only capsid mutant whose binding to RANCyp is affected by the isolated Q103E mutation. These patterns of host- and viral-specific interactions are suggestive of RANCyp being a barrier to cross-species transmission. While the order with which these mutations arose in the gorilla lineage is unknown, it is clear that they had the potential to inhibit their utilization by lentiviral capsids.

DISCUSSION

The cyclophilin domain of RANBP2 has been shown to be the main determinant of the RANBP2-capsid interaction (150). Here, we show that divergence in primate *RANCyp* affects interactions with a diverse set of capsid mutants that are representative of extant lentiviruses. Although the patterns of RANBP2 utilization by primate lentiviruses has yet to be established, the data presented here suggest a highly dynamic process, whereby interactions between RANCyp and capsid have been continuously

made and broken over evolutionary time. These observations, in combination with the evolutionary analysis that *RANCyp* has evolved under positive selection, are consistent with a scenario where the cylclophilin domain of RANBP2 is engaged in an arms race with capsid from primate lentiviruses.

In contrast to well-documented cases of arms races between HIV-1 and restriction factors (17, 19, 38, 85, 202), RANBP2 is an essential housekeeping gene. While there are now several documented cases of housekeeping genes rapidly evolving (24, 27, 138), the current study is the first to show that a component of the nuclear pore may be involved in an arms race. The nuclear pore and the cellular process of nuclear trafficking are both common targets of antagonism and utilization by viruses (255). However, it has been unclear how the host might counteract the hijacking of nuclear pore and trafficking proteins. Genes that encode proteins that constitute the nuclear pore or are involved in nuclear trafficking are generally well conserved in primates (data not shown), whereas only *RANBP2* and *Nup153* have been shown to be evolving under positive selection (27). Although the role of divergence in primate *Nup153* has yet to be explored, there is some overlap between positively selected sites and the regions known to mediate interactions with HIV-1 capsid (256).

This study reveals a strategy by which a nuclear pore protein, RANBP2, could avoid being utilized by lentiviral capsids. Specifically, the accumulation of amino acid-altering substitutions in *RANCyp* seems to force adaptation in the cyclophilin binding loop of lentiviral capsid. It is important to note that, because RANBP2 utilization by lentiviruses has not been broadly explored, it is possible that a given lentivirus might

adapt by bypassing the RANBP2 nuclear import pathway all together. This strategy is evident with the SIVmac lentivirus, which is cyclophilin-independent and does not bind RANCyp (Figure 5-3)(101, 226). Nevertheless, SIVmac does access the nucleus of non-dividing cells by a mechanism that has yet to be established. Therefore, while RANCyp and capsid seem to be involved in genetic conflict, some primate lentiviruses may completely step out of this conflict by adapting to use a RANBP2-independent nuclear import pathway.

The results shown in Figure 5-4 are consistent with lentivral capsids adapting to utilize RANCyp of a new species during cross-species transmission events. This would be indicative of *RANBP2* being a genetic barrier to cross-species transmissions of primate lentiviruses in nature. Most host genes that have been documented to be genetic barriers to transmission are either restriction factors of the innate immune system (28, 199) or cell surface receptors (25, 26, 89, 257). To determine with certainty that *RANBP2* is a *bona fide* genetic barrier would require more extensive analysis of *in vivo* infections of primates. Studies in the past have employed retrospective sequence analysis of primate center virus samples to show how SIVagm (African green monkey) adapted to rhesus macaque *TRIM5α* and *APOBEC3G* as it evolved into SIVmac. It will be impossible to find a similar dataset to study how SIVcpz adapted to gorilla RANBP2 as it evolved into SIVgor due to the heavy conservation restrictions on chimpanzees and gorillas. However, a cell-based assay where SIVcpz virus is passged on gorilla cells until it becomes well adapated is possible. The resulting viruses could then be sequenced and compared to the parental virus to look for adaptive mutations in the cyclophilin binding domain of capsid.

Until then, our data is only suggestive of *RANBP2* acting as a genetic barrier to cross-species transmission of primate lentiviruses.

To study the interaction between RANCyp and capsid, we utilized a TRIM-fusion assay that is becoming common for studying interactions with capsid (101, 158, 256, 258). By fusing the RANCyp domain downstream of the core domain of TRIM5α, we were able to use the restriction activity of this molecule as a proxy for interactions with capsid. However, there are several caveats to this method. First, these fusion molecules are located in the cytoplasm, which is distinct from the nuclear pore localization of full-length RANBP2 (259). Second, TRIM5α (and TRIM-fusion proteins) restricts retroviruses by forming a large hexameric lattice around the intact capsid core (260), whereas eight copies of RANBP2 extend out from the cytoplasmic face of the nuclear pore. Therefore, the architecture of the interaction between TRIM-RANCyp and capsid is very different than that of full-length RANBP2 and capsid. Third, It has been shown that a RANBP2 molecule that does not contain its cyclophilin domain will still bind capsid, although not as avidly (150). This suggests that there are regions of RANBP2 outside of the cyclophilin domain that are important for binding, and the TRIM-RANCyp assay used here, does not explore these regions. Another study even shows that the cyclophilin domain of RANBP2 is completely dispensible for HIV-1 replication (216). These conflicting reports will need to be reconciled in future studies, and the true role of the cyclophilin domain of RANBP2 remains to be fully characterized.

Nevertheless, there is evidence that the TRIM-RANCyp assay does faithfully predict interactions between full-length RANBP2 and capsid. In Figure 5-2, the assay

recapitulated the observation that SIVmac does not utilize human RANBP2 during infection. Another group has shown that lentiviral capsids that are sensitive to RANBP2 knockdown using siRNA are also susceptible to TRIM-RANCyp (224). An important observation of this study is that none of the lentiviral capsids that were sensitive to TRIM-RANCyp were shown to be insensitive to siRNA knockdown of RANBP2. Therefore, although experiments using full-length RANBP2 are certainly warranted, the TRIM-RANCyp assay is a convenient and accurate assay to use as a substitute. It is especially appealing considering that RANBP2 is over 3000 amino acids in length, which makes expression, purification, and biochemical assays using this protein particularly challenging.

The ability of primate lentiviruses to infect non-dividing cells by traversing the nuclear pore has been a long standing mystery. Only recently have molecular details come to light, which highlight capsid as a main determinant of nuclear access (210). Even more recently, specific nucleoporins, such a RANBP2, Nup214, Nup153, and Nup98, have all been shown to be involved in the nuclear import of the pre-integration complex of HIV-1 (150). We have now shown that the rapid evolution in the cyclophilin domain of RANBP2 affects interactions with lentiviral capsids in a species-specific manner, characteristic of a host-virus arms race and a genetic barrier to cross-species transmissions.

Chapter 6

Disruption of human-influenza evolutionary equilibrium reveals a novel restriction mechanism

Influenza A virus non-structural protein 1 (NS1) antagonizes the host interferon response through various mechanisms. One proposed mechanism involves binding the ubiquitin ligase TRIM25 and thereby inhibiting the ubiquitination of RIG-I and downstream signaling of the type I interferon response. However, the binding of TRIM25 by NS1s of various influenza strains is not perfectly correlated with inhibition of interferon signaling, suggesting that NS1 might bind TRIM25 to inhibit another antiviral function of TRIM25. Using a combination of evolutionary and functional analyses, we are able to show that TRIM25 binds directly to incoming influenza viral ribonucleoproteins in an RNA-dependent manner. This binding restricts influenza replication by inhibiting the transcription of viral RNA. The linker 2 domain of TRIM25 confers this antiviral activity and has been the target of recurrent positive selection during primate evolution, indicative of an ancient and ongoing genetic conflict between TRIM25 and influenza-like viruses.

INTRODUCTION

Influenza A viruse (IAV) causes annual epidemics that lead to severe respiratory disease in humans (261). Worldwide, millions of people are infected and hundreds of thousands succumb every year (262). The success of IAV is in no small part due to its ability to antagonize the host immune response. Specifically, one of IAV's eight RNA segments encodes non-structural protein 1 (NS1), which suppresses the innate immune system by shutting down host protein production and inhibiting the activation of interferon through various mechanisms (263, 264). One such mechanism is the binding and inhibition of the E3 ubiquitin ligase TRIM25 (264). TRIM25 is a member of the TRIM family of innate immunity proteins (135, 259, 265) and promotes the activation of the type I interferon response by ubiquitinating RIG-I, either covalently (266) or non-covalently (267). TRIM25 binding by NS1 inhibits the ubiquitination activity of TRIM25 and is thought to block the activation of the interferon response (264). However, some primary isolates of IAV do not inhibit the interferon response, even though their NS1 proteins bind TRIM25 (72). The lack of correlation between NS1-TRIM25 binding and interferon suppression led to the hypothesis that TRIM25 might have another function that is being antagonized by NS1 during IAV infection.

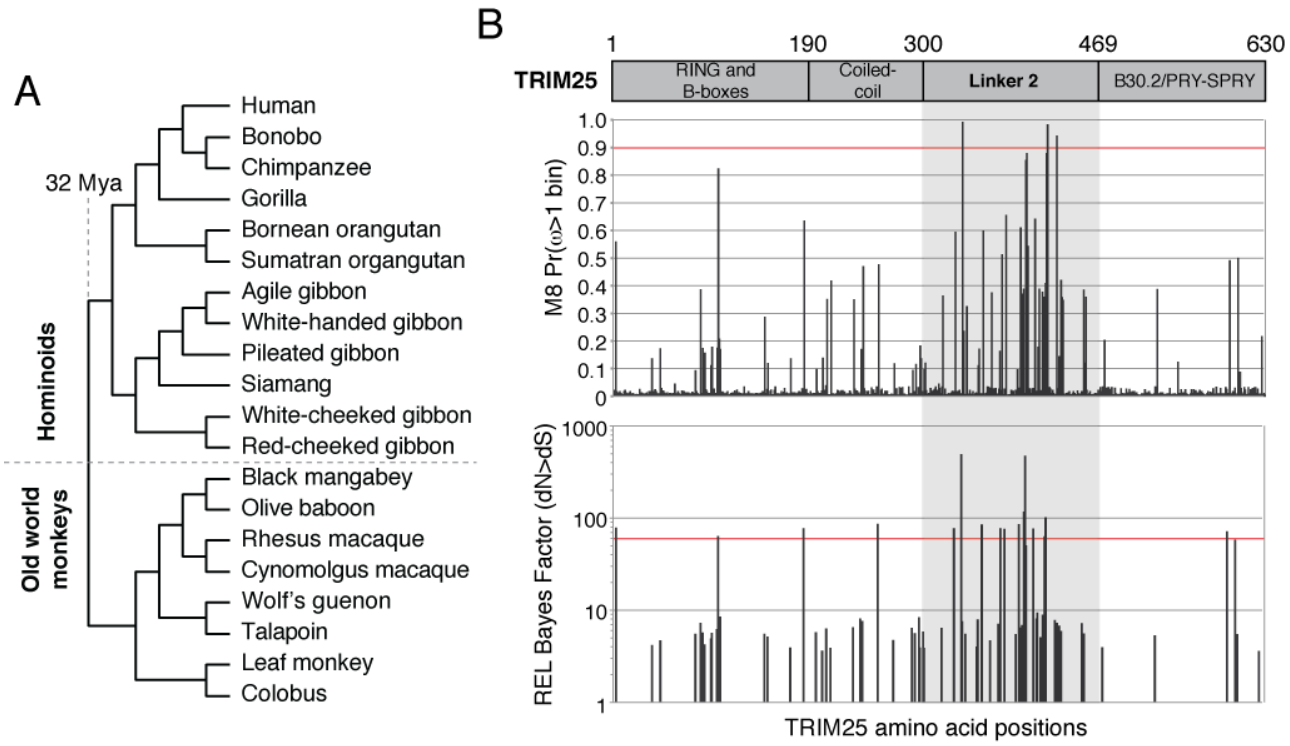


Table 6-1: Residues in TRIM25 identified using models of codon evolution

PAML M8[a]	REL[a]
Codons with dN/dS > 1[b,d] P > 0.9	Codons with dN/dS > 1[c,d] Bayes Factor > 50
H338, **L420**, **S429**	E3, N102, T185, T257, **H331**, **H338**, **L358**, **S376**, **V380**, **P394**, **L399**, **P400**, **A408**, **G419**, **L420**, K596, R604

[a] Dataset is a codon alignment of TRIM25 from the 20 simian primate species in (A).
[b] Codons assigned to the dN/dS >1 class in M8 with a posterior probability P > 0.90 by bayes empirical bayes. M8 is a significantly better fit than the null model, M8a ($p < 0.006$). Codon coordinates and the corresponding amino acid correspond to the human protein.
[c] Codons with a bayes factor > 50 for having dN > dS. Codon coordinates and the corresponding amino acid correspond to the human protein.
[d] Codon positions shown in bold are located in the Linker 2 domain.

Figure 6-1: *TRIM25* is under positive selection in primates. (A) Cladogram of the primate species represented in the *TRIM25* evolutionary analysis. **(B)** Domain diagram of TRIM25. The amino acid positions at domain boundaries are shown. Relevant statistics for individual amino acid positions are shown for the M8 model in PAML and Relative Effects Likelihood (REL). Red lines indicate significance cutoffs. Gray box highlights the Linker 2 domain.

RESULTS

Further evidence of a novel TRIM25 function is provided by the evolutionary history of TRIM25. TRIM25 was sequenced from a panel of 20 simian primates and found to be evolving rapidly under positive selection (Figure 6-1 and Table 6-1). Most amino acid residues experiencing positive are in the linker 2 (L2) domain of TRIM25, distinct from the coiled-coil domain that is known to mediate interactions with NS1 (268) (Figure 6-1 and Table 6-1). This suggests that amino acids in the L2 domain of TRIM25 have been rapidly selected for adaptation. Currently, the L2 domain has only been implicated in RNA binding, but it was not shown to be sufficient (269). The observation that TRIM25 is under positive selection in primates is in agreement with a previous study carried out on a more limited dataset (270). The evolutionary signature of positive selection in primates can highlight specific regions of proteins that interact with viruses and, in some cases, even be used to discover novel functions of previously uncharacterized protein domains (87, 126, 127).

Studies that use primate alleles to probe host-virus interactions have led to molecular insights about how viruses adapt to a host's innate immune response (51, 87, 88, 271). More importantly, they have highlighted the necessity to use nonhuman primate alleles when studying human viruses (133). Because human viruses have adapted to replicate in a human cell, they have developed mechanisms to bypass human antiviral factors. Therefore, by challenging a virus with nonhuman primate alleles, evolutionary equilibrium can be disrupted, and potent antiviral activity can be revealed.

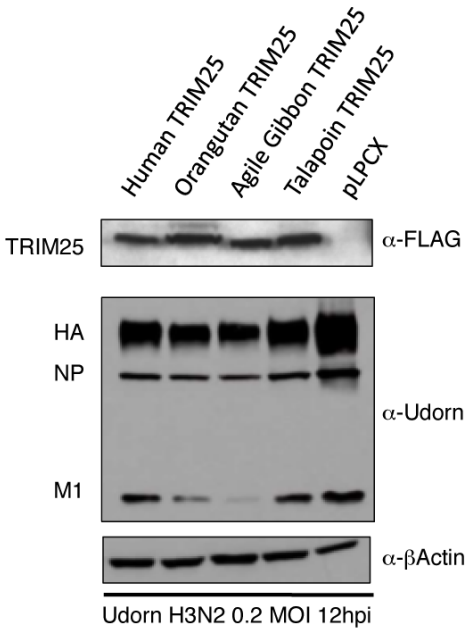


Figure 6-2: Primate TRIM25 alleles restrict influenza A virus protein levels. CRFK cells were used to stably express different FLAG-tagged TRIM25 alleles. Each cell line was infected at 0.1 MOI with the Udorn H3N2 influenza A virus. Whole-cell protein extracts were collected at 12 hours post infection and probed for the presence of the indicated proteins. A β–actin blot is shown as a loading control.

This strategy was used with TRIM25 to determine if its evolutionary history in primates has impacted interactions with IAV. Cell lines that stably express FLAG-tagged human, orangutan, agile gibbon, or talapoin TRIM25 were generated, along with an empty control cell line (Figure 6-2). Each of these cell lines was then subjected to a single-cycle infection using the Udorn H3N2 strain of IAV at 0.1 MOI. At 12 hours post infection, whole-cell lysate was obtained and probed for the presence of various influenza proteins using a Udorn-specific antibody that recognizes hemagglutinin (HA), nucleoprotein (NP), and matrix protein 1 (M1) (Figure 6-2). Primate TRIM25 alleles

restrict IAV protein levels to varying degrees. Therefore, genetic divergence in primate TRIM25 does affect antiviral activity against IAV.

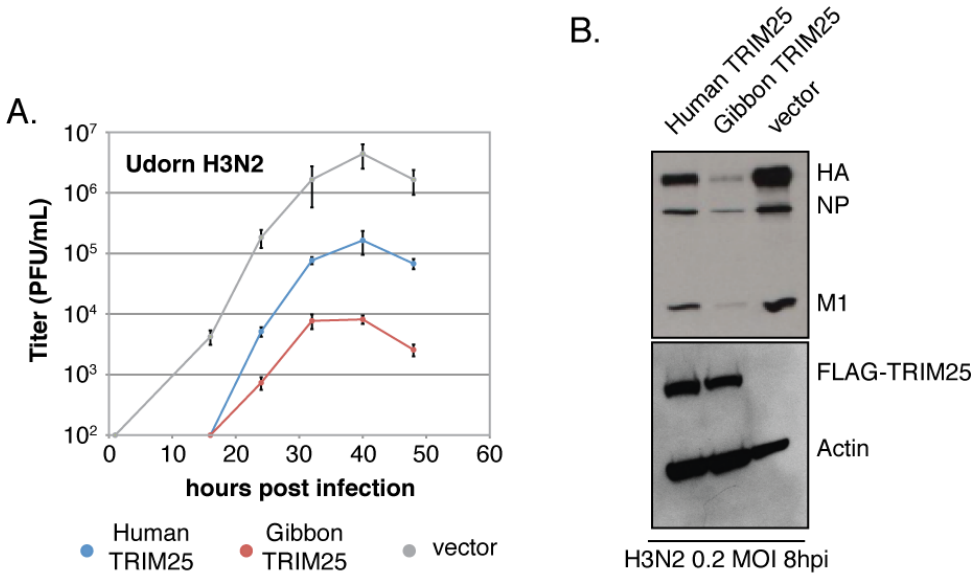


Figure 6-3: Human and agile gibbon TRIM25 restrict influenza A virus replication. CRFK cells stably expressing the indicated TRIM25 allele were infected with Udorn H3N2 influenza A virus at 0.02 MOI. Supernatant was collected at the indicated time points and titers were measured via plaque assay on an MDCK reporter cell line. Each growth curve was performed in triplicate and error bars represent the standard deviation of the mean. **(B)** Whole cell lysates from a single-cycle infection of CRFK cells stably expressing the indicated construct were probed for influenza virus proteins.

Next the antiviral activity of TRIM25 was tested during the full replication cycle of IAV. We focus on only the human and agile gibbon TRIM25 cell lines due to their large phenotypic differences (Figure 6-3). During a multiple-cycle infection, both the human and agile gibbon TRIM25 cell lines restrict IAV replication, although to different extents. Human TRIM25 inhibits IAV replication by more than a log, whereas agile gibbon TRIM25 inhibits IAV replication by more than two logs. Protein levels at an earlier time point also were determined (compare Figure 6-2 with

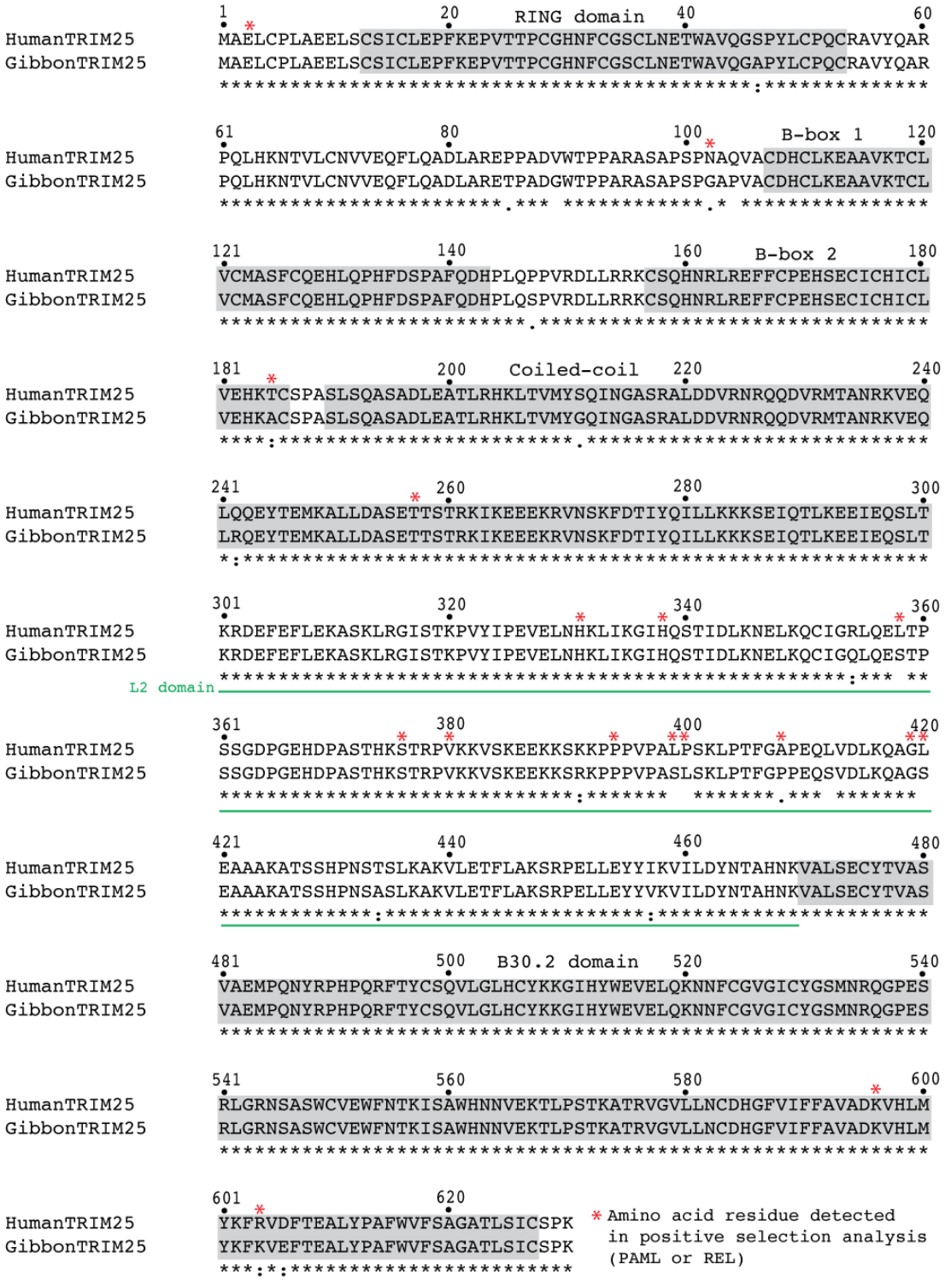


Figure 6-4: Protein alignment of human and gibbon TRIM25.

6-3B), and the amount of restriction was further exaggerated and in agreement with results from the growth curves. The human and agile gibbon TRIM25 proteins differ at only 21 out of 630 amino acid positions (Figure 6-4), suggesting that the increased antiviral activity of agile gibbon TRIM25 is conferred by a small number of genetic changes.

To characterize the mechanism by which agile gibbon TRIM25 restricts flu replication more potently than human TRIM25, we investigated the extent of interferon induction in the TRIM25 stable cell lines and also tested binding to IAV's NS1 protein. To test interferon induction, quantitative reverse transcription PCR was used to probe mRNA levels of IFN-β, Mx1, and ISG15 mRNA in non-infected cells and also cells infected at 2 MOI, 6 or 12 hours post infection (Figure 6-5A). Negligible differences between TRIM25 and control cell lines were observed with respect to interferon induction, suggesting that TRIM25 restriction activity is independent of its role in the RIG-I signaling pathway. Both human and agile gibbon TRIM25 bind NS1 with similar affinity, which shows that the increased antiviral activity of agile gibbon TRIM25 is not due to an evasion of NS1 binding (Figure 6-5B). Together, these results indicate that TRIM25 restriction activity is not mediated by any known function of TRIM25.

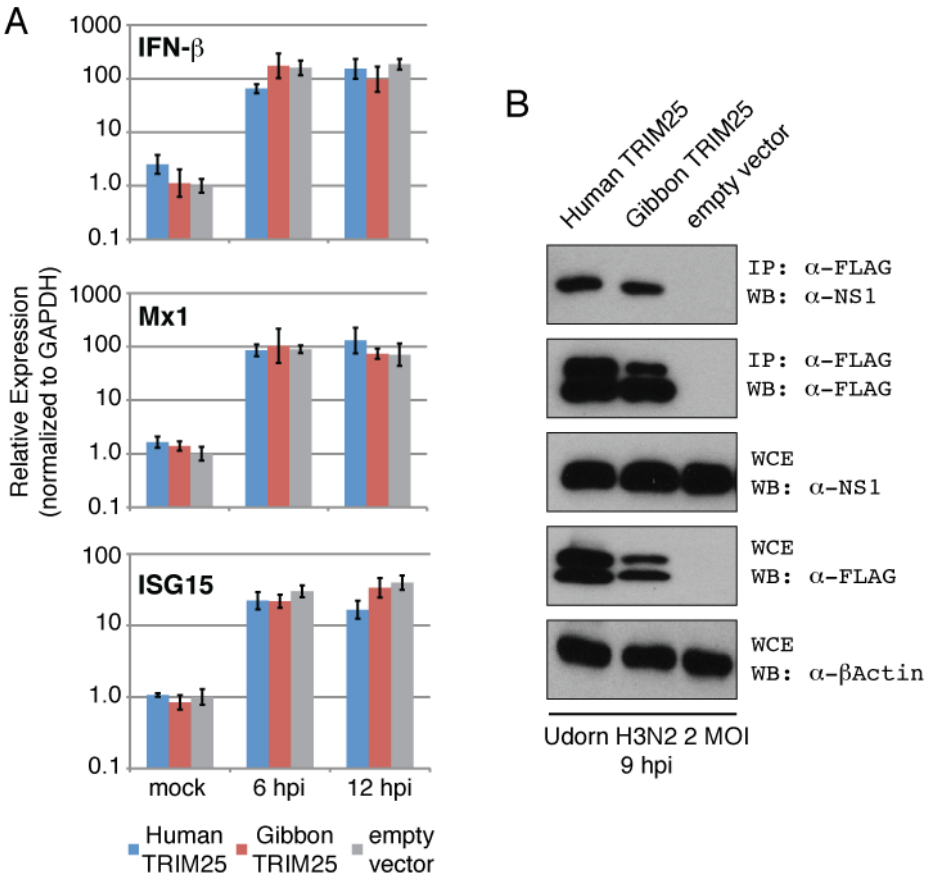


Figure 6-5: TRIM25 antiviral activity is not due to by interferon induction or binding to NS1. (A) Quantitate reverse transcription PCR was performed in CRFK cell lines stably expressing the indicated TRIM25 allele or empty vector. Cells were either mock infected or infected with Udorn H3N2 influenza A virus and harvested for total RNA at the indicated time points. cDNA was generated using oligo(dT) primers and probed using primers designed to detect the indicated feline gene. **(B)** Co-immunoprecipitation assay was performed by transiently transfecting the indicated construct in 293T cells and 48 hours later infecting with Udorn H3N2 influenza A virus. Whole cell protein lysates were collected 9 hours post infection.

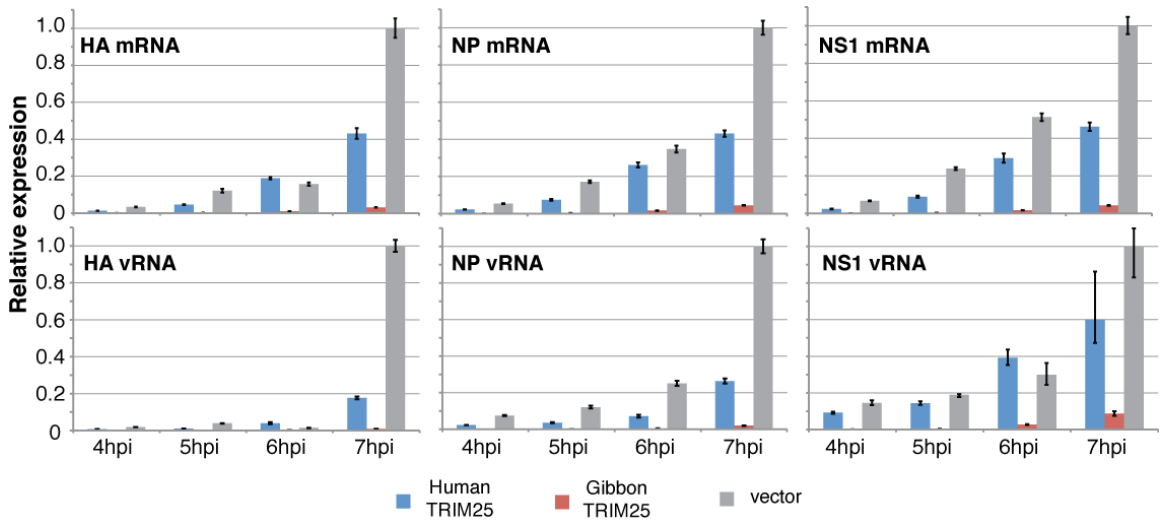


Figure 6-6: TRIM25 inhibits early transcription of both viral genomic RNA and mRNA. CRFK cells stably expressing the indicated TRIM25 construct or empty vector were infected with Udorn H3N2 influenza A virus at an MOI=2. Total RNA was isolated at the indicated time points and used to generate cDNA with either oligo(dT) primers for mRNA or viral specific primers for vRNA. The resulting cDNA libraries were then subjected to qPCR using primers for HA, NP, or NS1.

To gain further temporal resolution into TRIM25 restriction of IAV, early events in the influenza lifecycle were measured. Total RNA from infected cells at various times was isolated shortly after a single-cycle infection (4, 5, 6, and 7 hours post infection). These samples were then probed for viral genomic RNA and mRNA (Figure 6-6) using qPCR with cDNA templates generated using either oligo(dT) for mRNA or viral-specific primers for genomic RNA. The early time points probed in this experiment coincide with the first detectable transcription events during the IAV life cycle, shortly after the viral ribonucleoproteins have entered the nucleus (272). RNA levels of the HA, NP, and NS1 RNA segments, which represent structural and non-structural genes along with early and late genes, were measured in case there were any segment-specific effects. Both vRNA

and mRNA expression of all three segments was reduced in cell lines expressing TRIM25. In accordance with the growth curves shown previously (Figure 6-3A), agile gibbon TRIM25 exhibits greater antiviral activity than human TRIM25. Even at the latest time point, only a small amount of vRNA and mRNA are present, verifying the drastic differences in protein levels observed in Figure 6-3B. Therefore, TRIM25 restriction activity functions at or before the initial transcription of influenza genomic segments.

TRIM25 is expressed in the cytoplasm where it activates the RIG-I signaling pathway (266). The cytoplasmic localization of both human and agile gibbon TRIM25 was confirmed (data not shown). During the early stages of IAV infection, the only viral components that are exposed to the cytoplasm are M1 and the viral ribonucleoprotein (vRNP) (261), which is composed of three polymerase subunits (PA, PB1, and PB2) that are attached to one end of a double helical NP structure (273). The RNA segments of IAV wrap around the helical structure and form a loop at the end of the vRNP opposite to the polymerase complex. Because of the recently characterized RNA-binding property of TRIM25 (269), its antiviral activity against IAV could be due to vRNP binding mediated by RNA.

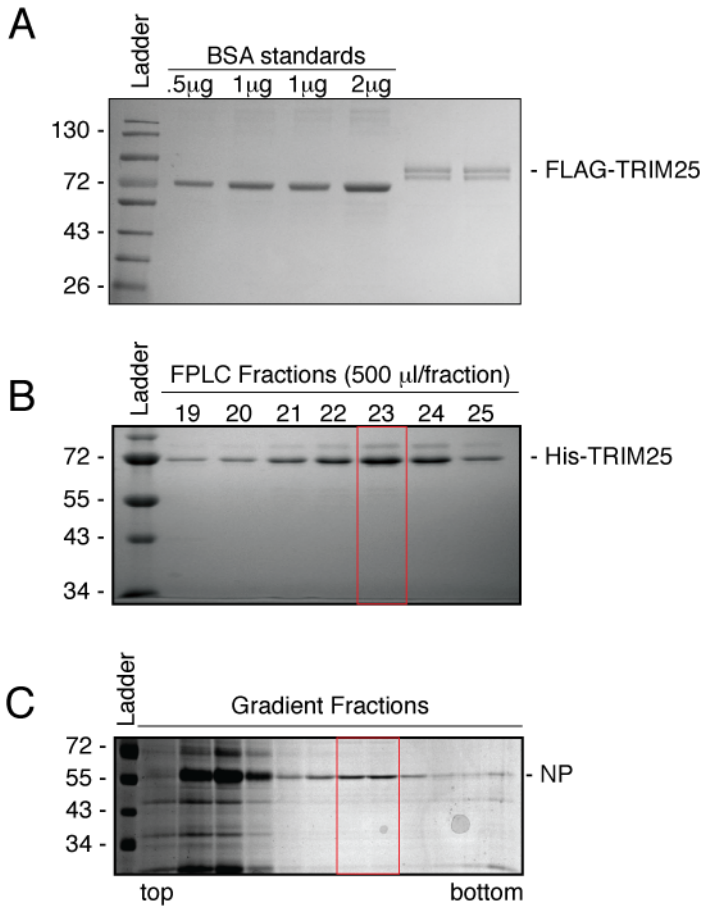


Figure 6-7: Purification of TRIM25 and IAV vRNPs. (A) Coomassie stained SDS-PAGE gel of FLAG-purified TRIM25. Whole-cell lysates from 293T cells transfected with FLAG-TRIM25 were used for a FLAG immunoprecipitation. **(B)** Coomassie stained SDS-PAGE gel of baculovirus expressed, His-tagged TRIM25 purified using size-exclusion FPLC. **(C)** Coomassie stained SDS-PAGE gel of purified IAV vRNPs. Red boxes in **(B,C)** represent fractions that were used in biochemical assays.

To test if TRIM25 binds IAV vRNPs, we purified human TRIM25 using FLAG immunoprecipitation from 293T whole cell extracts (Figure 6-7A) or a baculovirus expression system coupled with size-exclusion fast performance liquid chromatography (Figure 6-7B). IAV vRNPs were purified using supernatant from IAV infected canine

cells. Briefly, virions were pelleted and purified using sucrose cushions, disrupted using a detergent treatment to release vRNPs, and the vRNPs were further purified using a glycerol gradient (Figure 6-7C). Purified vRNPs and FLAG-TRIM25 were incubated together and then fractionated through a glycerol gradient. Fractions were collected and probed for the presence of TRIM25 and NP (Figure 6-8A). In the absence of vRNPs, TRIM25 remains in the top fractions. However, in the presence of vRNPs, TRIM25 migrates to lower fractions that coincide with the presence of NP. This indicates that TRIM25 is directly binding vRNPs. To test the role of RNA in TRIM25 binding to vRNPs, we pretreated vRNPs with RNase prior to incubation with TRIM25. This resulted in a loss of co-migration between TRIM25 and vRNPs. Therefore, TRIM25 binding to IAV vRNPs is RNA-dependent. To investigate where on the vRNPs TRIM25 is binding, we used electron microscopy to visualize TRIM25-vRNP complexes. His-tagged TRIM25 was used and conjugated to Ni-NTA-nanogold beads. Unlabeled vRNPs can be visualized and were confirmed to be in their native conformation (Figure 6-8B). Both human and gibbon TRIM25 bound to vRNPs, although we observe more nanogold particles bound to vRNPs when using gibbon TRIM25 (Figure 6-8B). In the case of gibbon TRIM25, the entire length of the vRNP appears to be bound. However, the majority of bound vRNPs are bound at the ends, suggesting that TRIM25 initially binds at the polymerase complex or the RNA loop.

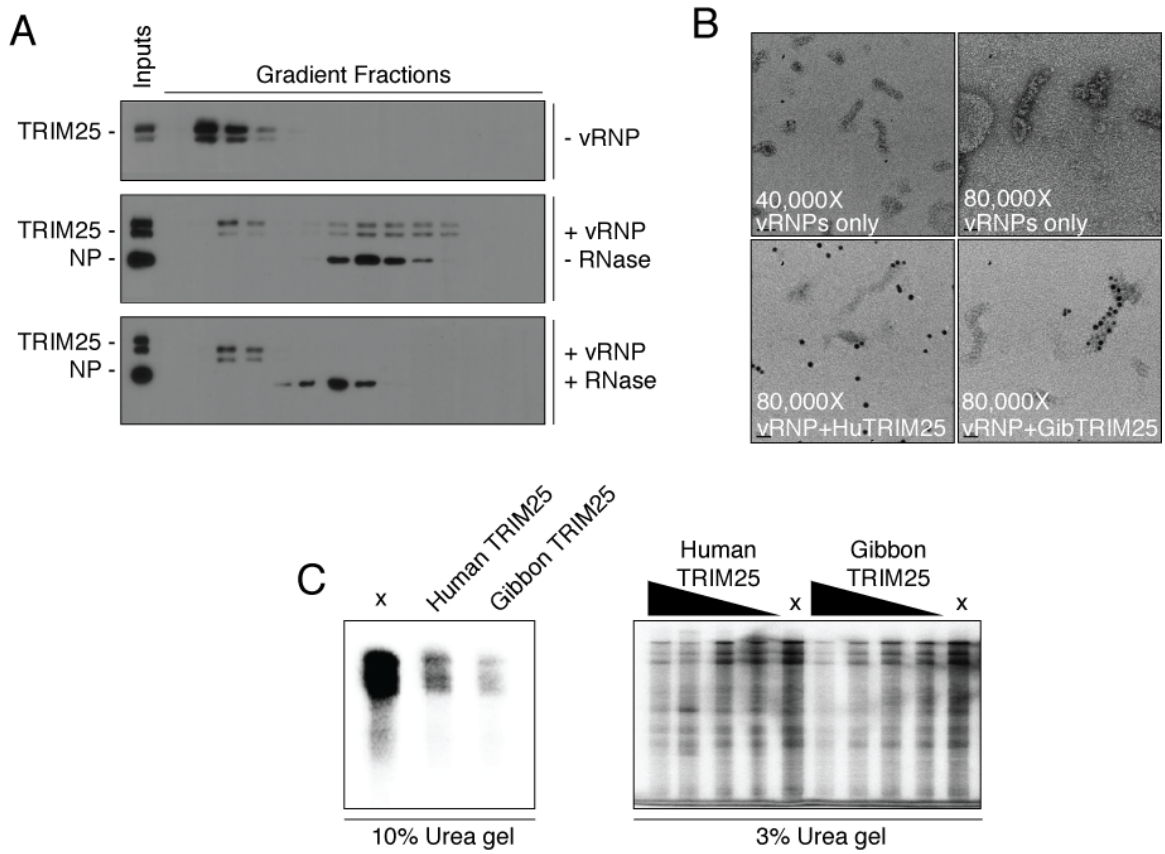


Figure 6-8: TRIM25 binds IAV vRNPs and inhibits transcription. (A) Western blots of glycerol gradient fractions from TRIM25 and vRNPs incubations. Glycerol concentration increases from left to right. **(B)** Electron microscopy images of vRNPs and Ni-NTA-nanogold labeled TRIM25. **(C)** vRNP transcription assay. Purified vRNPs were mixed with cellular mRNAs and labeled rNTPs. Products were separated on a urea gel and imaged using a phosphoscreen.

To connect TRIM25-vRNP binding to TRIM25 antiviral activity, this binding was assessed for effects on vRNP function. Transcription activity of purified vRNPs can be monitored by mixing them with cellular mRNAs and ribonucleotide triphosphates. Radiolabelled uracil is used so that transcripts can be visualized (Figure 6-8C, far left lane). When human or gibbon TRIM25 is pre-incubated with vRNPs, transcription activity is significantly reduced (Figure-6-8C). Again, gibbon TRIM25 was observed to

have more antiviral activity than human TRIM25, and this effect can be titrated with increasing amounts of TRIM25 to a point where viral transcripts are almost undetectable (Figure 6-8C, right gel). Therefore, TRIM25 binding to IAV vRNPs blocks transcription by the viral polymerase complex.

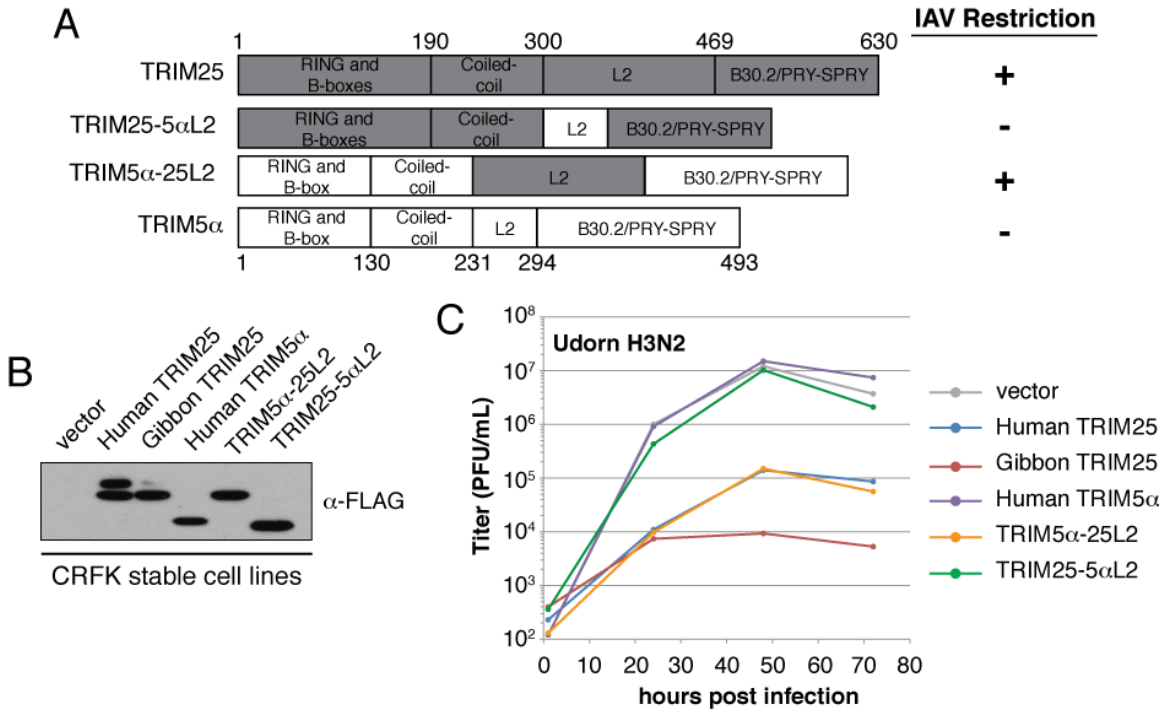


Figure 6-9: The L2 domain of TRIM25 mediates restriction activity against IAV. **(A)** Domain diagram schematic of TRIM25 and TRIM5α chimeras. Domain boundaries for TRIM25 and TRIM5α are provided as amino acid coordinates above and below, respectively. Each construct is drawn to scale. **(B)** Western blot of whole-cell lysates from feline cells that stably express the indicated FLAG-tagged construct. **(C)** IAV infection of stable cell lines expressing wild-type and chimeric alleles. Cells were infected at 0.002 MOI and supernatants were collected at 1, 24, 48, and 72 hours post infection. Canine reporter cells were infected to calculate titers using plaque assays.

The L2 domain of TRIM25 has rapidly diverged during the course of primate evolution (Figure 6-1). To determine the contribution of this domain to IAV restriction, L2 chimeras were generated between TRIM25 and a closely related TRIM protein, TRIM5α (Figure 6-9A). TRIM proteins are modular and chimeras between different TRIMs have been utilized previously to determine domain function (209). Stable cell lines expressing TRIM25/TRIM5α chimeras were generated (Figure 6-9B), infected with IAV under multiple-cycle growth conditions, and titered on reporter cell lines (Figure 6-9C). Human and gibbon TRIM25 restrict IAV to similar levels as before (compare Figure 6-3A and Figure 6-9C), whereas TRIM5α does not restrict IAV. However, when the L2 domain from human TRIM25 is inserted into TRIM5α (TRIM5α-25L2, orange curve), this construct gains the ability to restrict IAV at a level similar to human TRIM25. Conversely, insertion of the L2 domain from TRIM5α into TRIM25 (TRIM25-5αL2, green curve) confers a near complete loss of restriction. Therefore, the L2 domain of TRIM25 is both necessary and sufficient for restriction of IAV.

DISCUSSION

Using a combination of evolutionary and functional analyses, a novel function of TRIM25 has been defined. By binding IAV vRNPs and inhibiting transcription of the viral polymerase complex, TRIM25 potently inhibits the replication of IAV. This is not the first example of a TRIM protein that antagonizes a viral polymerase (274, 275), however, it is the first example of direct binding to vRNPs. This mechanism most likely relies on the unique double helical, elongated structure of IAV vRNPs for two reasons.

First, no TRIM25 antiviral activity in a IAV mini-genome assay, where all components of the vRNPs (PA, PB1, PB2, NP, and vRNA) are present (data not shown). In these assays, fully formed vRNPs are not present in the cytoplasm and the state of the polymerase complex, due to overexpression, is most likely not representative of a live infection. Second, no interaction between TRIM25 and any single protein component of vRNPs was detectable in an immunoprecipitation assay where TRIM25 was co-transfected with PA, PB1, PB2, or NP (data not shown). These negative results support a model where TRIM25 binds a structural element of vRNPs that is not present in any single componenet of vRNPs.

This quality of TRIM25 is reminiscent of another TRIM protein, TRIM5α. By recognizing fully intact retroviral capsid cores (45), TRIM5α disrupts the retroviral lifecycle in the cytoplasm shortly after reverse transcription takes place (276). Unique molecular patterns like the retroviral core or the vRNPs of IAV are ideal targets for innate immune proteins because their structures must be maintained and are therefore evolutionarily conserved. However, in the case of TRIM5α, escape mutations in the capsid core have been observed, and these mutations have helped to identify the interface that TRIM5α binds (277). Future studies with TRIM25 should be aimed at identifying the interface of IAV vRNPs that TRIM25 recognizes.

This evolutionary analysis identified the L2 domain of TRIM25 as the determinant of IAV restriction. Because a primate dataset was used, this result suggests that primates have been antagonized by influenza or influenza-like viruses in the past. Although infectious IAV has not been isolated from nonhuman primates, some primates

make antibodies that are active against human IAV (278, 279). This suggests that nonhuman primates are being infected by IAV in the wild, which would mean that primate TRIM25 alleles are actively being selected. This scenario is in line with the evolutionary signature of positive selection that we have documented in primate TRIM25, and suggests that TRIM25 and IAV vRNPs have been engaged in a genetic conflict with each other for many millions of years.

MATERIALS AND METHODS

Primate biomaterials used

Primary and immortalized primate cell lines from primate species were grown in standard media supplemented with 15% fetal bovine serum at 37ºC and in 5% CO_2.

Primate *TRIM25* gene sequences

Human Refseq sequences for *TRIM25* were obtained from the NCBI nucleotide database. Chimpanzee, orangutan, rhesus macaque, and marmoset gene sequences were obtained from the UCSC genome database (http://genome.ucsc.edu/) using the BLAT alignment tool. Genes were sequenced from 15 additional primate species, and from chimpanzee, orangutan, rhesus and marmoset in instances where the genome-project sequences were of poor quality. PCR or RT-PCR was performed from total RNA, gDNA, or cDNA with SuperScript III One-Step RT-PCR system with Platinum Taq (Invitrogen, #12574-018), PCR SuperMix High Fidelity (Invitrogen, #10790-020), or Phusion High

Fidelity PCR Master Mix (NEB, #F-531S). Primers were designed in conserved regions of the 5' or 3' UTR to amplify full open reading frames.

PAML and REL analysis

Codon models were tested with codeml in the PAML 4.1 software package (93). To detect selection, multiple alignments were fit to the NSsites models M8a (neutral model, codon values of dN/dS fit to a beta distribution plus an extra codon class fixed at dN/dS = 1) and M8 (positive selection model, similar to M8a but with the extra class allowed to be dN/dS >1). A likelihood ratio test was performed to assess whether permitting codons to evolve under positive selection gives a significantly better fit to the data (model comparison M8a vs. M8). In situations where the null model could be rejected ($p < 0.05$), posterior probabilities were assigned to individual codons belonging to the class of codons with dN/dS > 1. REL analysis was performed using the datamonkey.org webserver (280). In this model, three classes each of dN and dS are estimated and a bayes factor for dN > dS (dN/dS > 1) is calculated.

TRIM25, TRIM5α, and TRIM25/TRIM5α chimera expression constructs

Primate TRIM25 constructs were amplified from primate cDNA. TRIM25/TRIM5α chimeras were constructed by generating fragments with 20-25bp overlapping regions of TRIM25 and TRIM5α. Overlapping fragments were spliced together using PCR with each fragment as a template and outside flanking primers. Constructs were TA-cloned into pCR4 (Invitrogen, #K4575-01). An N-terminal FLAG

tag was added by PCR and these tagged constructs were TA-cloned into the Gateway entry plasmid pCR8 (Invitrogen, #K2500-20). An LR Clonase II reaction (Invitrogen, #11791-100) was used to transfer these constructs into a Gateway-converted pLPCX retroviral vector (Clontech, # 631511).

Generation of stable cell lines

To generate cell lines that stably express primate TRIM25 and TRIM25/TRIM5α chimeras, retroviral vectors were used to transduce CRFK cells (ATCC). To generate the retroviral vectors, 293T cells were seeded at a concentration of $1x10^6$ cells/well in a 6-well dish. After 24 hours, each well was transfected with 2 μg pLPCX construct (empty or encoding the gene fragment of interest), 1 μg pCS2-mGP encoding MLV *gag-pol* (142), and 0.2 μg pC-VSV-G at a final 1:3 ratio of DNA to TransIT-293 (mg DNA : ml TransIT-293). Supernatants were collected after 48 hours, passed through a 0.2 μm filter, and used to infect CRFK cells grown in DMEM supplemented with 10% fetal bovine serum, 100 IU/ml penicillin, 100 μg/ml streptomycin, and 2 mM L-glutamine. After 24 hours, media containing 8 μg/ml puromycin was added to select for transduced cells. Expression of TRIM25 and TRIM5α alleles constructs was detected by Western blot.

Antibodies and Western blot analysis

Cell lines were grown to confluency in a 6-well dish, collected using a cell scraper, and lysed in a buffer containing 150 mM NaCl, 50 mM Tris-HCl (pH 7.4), 1% NP-40, and Complete protease inhibitor (Roche, #11836170001). After quantitation of

protein concentration using a Bradford assay, 30 μg of whole cell extract was resolved using a 10% polyacrylamide gel and transferred to a nitrocellulose membrane. Udorn-specific antibody was a kind gift from Robert Lamb. NS1-specific antibody was generated by injecting rabbits with purified GST-NS1 from Udorn virus. FLAG-tagged constructs were detected using a 1:2000 dilution of anti-FLAG antibody (Roche, SydLabs, #PA000274-P-210). β-actin was also analyzed as a loading control using a 1:1000 dilution of mouse anti-β-actin (Santa Cruz, #sc-47778). A 1:10,000 dilution of donkey mouse-specific horseradish peroxidase-conjugated antibody (Thermo Scientific, #32430), donkey rabbit-specific horseradish peroxidase-conjugated antibody (Thermo Scientific, #32460), or donkey goat-specific horseradish peroxidase-conjugated antibody (Santa Cruz, #2020) was used as a secondary probe. Blots were developed using the ECL Plus detection reagent (GE Healthcare, #RPN2132).

Influenza A virus infection

Single-cycle and multiple cycle influenza A virus infections (A/Udorn/H3N2) were performed at a MOI=2. CRFK cells were seeded in a 6-well dish at $8x10^5$ cells/well. After plating for 24 hours, cells were washed with PBS and then incubated in infection media (DMEM supplemented with Pen/Strep, L-Glut, and 1% BSA) for one hour at 37°C. Cells were washed once more in PBS and incubated in influenza growth media (DMEM supplemented with Pen/Strep and L-Glut; multiple-cycle samples also included 1.0 μg/ml N-acetylated trypsin). For single-cycle infections, cell lysates were harvested at 8 hours post infection using RIPA buffer supplemented with complete protease inhibitor

(Roche) and PMSF (Invitrogen). For multiple-cycle infections, each cell line was seeded in triplicate and infected with a unique dilution of influenza A virus. Supernatants were collected at 1, 12, 24, 48, and 72 hours post infection and titered by a plaque assay on MDCK reporter cells. Infections for plaque assays were performed as described above, except the final incubation media contained 1.4% Avicel (Sigma) to induce plaque formation.

Immunoprecipitation

293T cells were seeded at $1x10^6$ cells/well in a 6-well dish. After 24 hours, cells were transfected with 2 μg of pLPCX-TRIM25 or empty vector using TransIT-293 (Mirius Bio, #MIR2704) at a 1:3 μg DNA:reagent ratio. After 48 hours of transfection, cells were infected at 2 MOI with Udorn influenza A virus. At 9 hours post infection, cells were lysed in 250 μl co-IP buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA, 2.5 mM $MgCl_2$, 1% NP-40, 10% glycerol). Whole-cell lysate (30 μl) was saved for inputs and the remainder was diluted 1:1 with Tris-HCl/NaCl buffer and pre-cleared using Protein G Dynabeads (Life Technologies, #10003D). Beads were added to WCE and rotated at 4° for one hour. Beads were then removed and 1.5 μl of FLAG-specific antibody was added to the supernatant and rotated overnight at 4°. The next day, beads were washed three times for ten minutes in IP wash buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.25% NP-40) at 4°, and then eluted in 30 μl FLAG elution buffer (50 mM Tris-HCL pH 7.4, 150 mM NaCl, 0.012% Triton-X, 10% glycerol, 0.2 mM EDTA, 500

ng/mL 3xFLAG peptide (Sigma Aldrich, #F4799)). Eluates were separated using SDS-PAGE electrophoresis and visualized using a Western blot protocol.

Quantitative reverse transcription PCR

Templates for qRT-PCR were generated by infecting target cells with Udorn influenza A virus at MOI=2 and collecting RNA at 4, 5, 6, and 7 hours post infection using the RNeasy Mini Kit (Qiagen, #74104). RNA (500 ng) was then used to generate cDNA using either oligo(dT) or virus-specific primers with Superscript III First-Strand Synthesis System (Invitrogen, #18080-051). cDNA was diluted 1:25 and used as a template in qPCR with Power SYBR Green PCR Master Mix (Applied Biosystems, #4367659) and virus- or host-specific primers (see Figures 6-6 and 6-7). Cycling for qPCR was performed in Life Technologies' Vii-A7 Real-Time PCR System. All samples were normalized to GAPDH levels.

Purification of TRIM25

FLAG-tagged TRIM25 was purified as described in the immunoprecipitation protocol. For His purification, TRIM25 open reading frame was inserted into pFastBac-HTa (Invitrogen, #10712-024) vector downstream of a 6xHis tag using EcoRI and XhoI restriction sites. This plasmid was then transformed into DH10Bac cells (Invitrogen, #10359-016) to induce recombination of His-TRIM25 into a Bacmid. Bacmids were isolated and transfected into insect cells to generate baculovirus that harbors His-TRIM25. Sixty 15cm plates of insect cells were then infected with TRIM25-harboring

baculovirus, and cells were collected 48 hours later. Cells were lysed in TRIM buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 50 mM $ZnSO_4$, 10% glycerol, 100 mM 2-βMercaptoethanol), dounced and sonicated, and centrifuged to pellet cell debris. One mL of 50% Ni-NTA slurry per 4 mL of cleared lysate was added and rotated at 4° for 1-2 hours. Beads were collected and washed (50 mM Tris-HCl, 150 mM NaCl, 20 mM imidazole, 0.5% NP-40) three times. Beads were eluted using wash buffer with 250 mM imidazole. Eluates were subjected to fast performance liquid chromatography using a Superdex 200 3.2/300 size exclusion column (GE Healthcare, #29-0362-32). Fractions with UV peaks at the appropriate size for TRIM25 were used in downstream analyses.

Purification of IAV vRNPs and binding experiments

Supernatant from Udorn-infected MDCK cells was clarified by centrifugation at 2600xg for 10 minutes at 4°. The clarified supernatant was then layered over 5 mL of 20% sucrose in buffer A (100 mM NaCl, 10 mM Tris-HCl pH 7.4, 2 mM EDTA). The virus was pelleted by ultracentrifugation at 104,000 xg for 45 minutes at 4° in a SW28 rotor. The virus pellet at the bottom was then resuspended in buffer A. Further viral purification was achieved by adding the resuspended pellet to a linear sucrose gradient (30-50% w/w) in buffer A and separated at 112,600xg for 3 hours at 4° in a SW41 Ti rotor. The banded virus was collected and then disrupted to release vRNPs using disruption buffer (5 mM $MgCl_2$, 100 mM KCl, 1.5 mM dithiothreitol, 5% glycerol, 10 mM Tris-HCl pH 8.0, 1% Triton-X 100, 10mg/mL lysolecithin). To purify the vRNPs, a discontinuous glycerol gradient (1 mL 70%, 0.75 mL 50%, 0.375 mL 40%, 1.8 mL 33%,

all v/v) is prepared in an ultraclear centrifuge tube (Beckman, 13mm x 51mm). The sample was then centrifuged at 217,000 xg for 3 hours and 45 minutes at 4° in a SW55 Ti rotor. Fractions were then taken from the gradient and analyzed for the preseence of nucleoprotein (NP). Fractions containing mainly NP were pooled and used for downstream analyses.

For TRIM25-vRNP binding, vRNP samples were dialysed with TRIM25 wash buffer and then incubated with His-TRIM25 at 4° for 1-2 hours. This sample was then loaded onto a discontinuous glycerol gradient and ultracentrifuged as described above. Fractions were probed for the presence of NP or TRIM25.

Electron microscopy

Purified vRNP and TRIM25 samples were mixed at 1:1 NP-to-Trim25 mass ratio and incubated for 1h at 4°C. The mixture (5 μl) was then added onto glow-discharged FCF400-Cu grid (Electron Microscopy Sciences). After 1 min, extra liquid on the sample grid was removed by gently blotting the edge of the grid on a filter paper. The grid was then placed upside-down on a droplet of 5 nm Ni-NTA-Nanogold (Nanoprobes) diluted 5-fold with 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 5 mM imidazole, and incubated for 30 minutes at room temperature. Afterwards, the grid was washed twice with 5 mM imidazole and rinsed twice with water before staining with freshly prepared 0.75% uranyl formate solution for 90 seconds. After air-drying overnight, the grids were examined using JEOL 1230 High Contrast Transmission Electron Microscope at 80kV.

In vitro transcription of IAV vRNPs

Transcription using purified vRNPs was conducted as described previously (281). Briefly, purified vRNPs were incubated with or without TRIM25 (see above) and then cellular mRNAs were added to provide cap-snatching substrate for vRNPs. Labeled ribonucleoside triphosphates were used to polymerize transcription and visualize transcription products. Transcription reactions were resolved on urea-PAGE gels and transcription products were visualized using a phosphoscreen and phosphoscanner.

Chapter 7

Concluding remarks and future direction

The practice of applying the evolutionary arms race concept to molecular host-virus interactions in primates is still in its infancy. Even amongst mammals, only a few well-documented cases exist. It has only recently been appreciated that the evolutionary signature of positive selection can help to elucidate novel antiviral function, the rules of engagement for host-virus interactions, and genetic barriers to cross-species transmission of viruses. In the following chapter I will elaborate on how my own studies have advanced those topics and propose how evolutionary thinking can continue to benefit the study of host-virus interactions. There are also several unexpected applications of host-virus arms races that will be the core of my future studies. From hunting for new model systems to designing potent restriction factors from scratch, these new directions will continue to build on their foundation—a signature of positive Darwinian selection discovered in DNA.

Host-virus arms races are everywhere, but you must look closely

Viruses have exerted evolutionary pressure on their hosts throughout the history of life on earth. Specifically, viruses have affected mammalian genomes by forcing innovative immune strategies and even by depositing themselves in their host's DNA (124, 125, 282). Also, hosts will sometimes repurpose viral genes for their own good (283). These observations provide evidence that most mammalian virus families have ancient origins dating back millions, in some cases hundreds of millions, of years (284, 285). As viruses and hosts interact with each other over these long spans of time, relics of these battles are left in the genomic content of each organism involved. In the previous chapters I have described several cases where I have uncovered these relics using statistical analyses that detect the genetic signature of positive selection. These genomic signatures, in the context of genes involved in virus replication, are indicative of a host-virus arms race and can be used to elucidate details of ancient interactions.

The most well-characterized host-virus arms races come from the HIV-1 field. Since the discovery of the restriction factor TRIM5α (286), and the subsequent characterization of its evolutionary history in primates (17, 55), innate immunity proteins have been prime candidates for genetic conflict with viruses. Although there have been four other HIV-1 restriction factors that have been shown to have a dynamic evolutionary past (19, 38, 84, 85, 287), evolutionary characterization of restriction factors against other viruses is lacking. Aside from retroviruses, there are only concrete examples for poxviruses (51, 271), thogotoviruses (87), and hepaciviruses (88). With so many more

diverse families of mammalian viruses, it is likely that many more host-virus arms races will be revealed in the near future.

In Chapter 6, I characterize the restriction activity of TRIM25 against influenza A virus and show that the Linker 2 domain of TRIM25 mediates this restriction. I also show that the linker 2 domain has been the target of positive selection during primate evolution. Although the exact molecular interface that TRIM25 binds is yet to be determined, the interaction between TRIM25 and flu has likely been highly dynamic over evolutionary time. This would be the first example of a host-virus arms race between flu and an innate immunity protein. However, there are most likely many more to be uncovered because several other proteins have been shown to have restriction activity against influenza (39, 288, 289). Future studies should aim at characterizing the evolutionary history of the genes that encode these antiviral factors. More generally, restriction factors against any other virus are prime candidates for genes involved in host-virus arms races.

Aside from restriction factors, many other host proteins that interact with viruses could potentially be engaged in host-virus arms races. In Chapter 3, I show how some genes that *promote* HIV-1 replication can also evolve under positive selection (27). What is interesting about these cases is that these genes tend to have critical housekeeping functions, such as microtubule assembly, nuclear pore trafficking, or DNA repair (24). Nevertheless, they have found a way to maintain housekeeping function while simultaneously engaging in a host-virus arms race with a virus. This can sometimes be achieved because the region of the gene that is critical for housekeeping function is

distinct from the molecular interface that interacts with viruses (26). Therefore, absolutely any host-virus interaction should be considered as a candidate for a host-virus arms race. The possibilities are staggering, given that a virus such as HIV-1 interacts with over a thousand proteins (76, 115-118). Similarly, genome-wide siRNA screens for other viruses are revealing massive amounts of host proteins that are involved in viral life cycles (291-293). Most host genes that interact with viruses will be evolutionarily constrained due to their housekeeping function, however, a subset will have the flexibility to engage in an arms race, and I have provided a screening method for identifying such genes in Chapter 3.

My studies have focused on primate genes, however, there are instances where other animal clades are more appropriate for evolutionary analysis. For example, arenaviruses are endemic in the rodent clade and occasionally zoonosis occurs that leads to hemorrhagic fever in humans (294). Because this virus spends most of its time in rodents, evolutionary signatures of positive selection that relate to arenavirus infection can be found in a dataset consisting of rodent genes but not primate genes (26). A similar story has played out with respect to SARS-CoV in bats (25). Therefore, to properly conduct an investigation into a potential host-virus arms race, one must not only have a candidate gene in mind, but also a candidate animal clade that relates to the ecology of the virus in question. In this respect, evolutionary signatures of positive selection can be used as evidence that a particular animal clade is serving as a reservoir for a virus (25). This is one of the biggest surprises that has stemmed from such analyses, and is currently being used to search for the Ebola reservoir.

Evidence is now emerging that provides a connection between genes involved in host-virus arms races and genetic barriers to cross-species transmission of viruses. The amount of work that is required to show this definitively is staggering, nevertheless, it has now been shown for *TRIM5α* and *APOBEC3G* (28, 199). This group was able to retrospectively analyze samples taken from a primate center during a cross-species transmission event of simian immunodeficiency virus and show that the evolution of the virus coincided exactly with evasion of *TRIM5α* and *APOBEC3G* genotypes. In Chapter 5, I show that interactions between RANBP2, a molecule involved in nucleocytoplasmic trafficking, and lentiviral capsids is species-specific and correlates to cross-species transmission events. This is the foundation of a case for RANBP2 being a barrier to cross-species transmissions of primate lentiviruses. Further evidence will come from studying how lentiviruses adapt to new RANBP2 alleles. If *in vivo* data is ever available for this interaction (my studies highlight gorilla and chimpanzee interactions, and both of these species are highly endangered) it would provide the greatest evidence. Until then, I can only speculate that RANBP2, a gatekeeper to nuclear access, is engaged in a host-virus arms race with lentiviral capsids that mediates cross-species transmissions of primate lentiviruses in nature.

Virology groups that typically do not employ evolutionary analyses are now using them as a standard to identify functional domains in host proteins that interact with viruses (101, 158, 287, 295), and this is perhaps the greatest evidence of their value. With the number of host-virus interactions growing exponentially as more genome-wide data

becomes available (296, 297), evolutionary thinking is likely to spread into every corner of virology research.

Bioprospecting for the future

In Chapter 4, I show how an amino acid residue in *CD4* that is evolving under positive selection, and is polymorphic within Spix's owl monkeys, mediates entry of HIV-1. These owl monkey *CD4* alleles are then shown to be permissive to primary isolates of HIV-1, a characteristic that is lacking in current model systems of HIV-1. I am now in the process of isolating CD4+ T-cells from this owl monkey species to test their permissivness for the entire HIV-1 lifecycle. If they are fully permissive, this species would be a viable candidate for an HIV-1 model system. This study highlights the significance of unchartered territory in model systems biology—genetic polymorphisms in nonhuman primate species (298-300).

It is well accepted that polymorphisms in human populations can affect the outcome of virus infections and diseases (37, 301, 302). However, the exploration of this concept in model organisms is lacking. Genetic loci that affect the outcome of viral infections are constantly being characterized, but rarely do these loci get genotyped in relevant model systems. As opposed to viewing this as a problem for existing model systems, I see it as an opportunity to discover individuals with unique genotypes in species that might have been prematurely discounted as model organisms.

Consider the case of the HIV-1 model organism. Currently, the major model organism for HIV-1 infection is the rhesus macaque (172, 303). Putting all socioeconmic

issues aside, this species is a poor model system because rhesus macaques do not encode a permissive receptor for HIV-1 (174), and they encode several restriction factors that are active against HIV-1 (172). To solve this issue, chimeric viruses have been generated that encode parts of HIV-1 and parts of simian immunodeficiency virus, the primate relative of HIV-1. However, these chimeric viruses are genetically distant enough from HIV-1 that vaccines raised against them are not relevant to primary isolates of HIV-1. This has been an incredible barrier to generating a viable vaccine against HIV-1, and even the pathogenesis of these chimeric viruses in rhesus macaques do not emulate the pathogenesis of HIV-1 in humans. And so it is perfectly plausible to pose the question, "Have we found the right model organism?"

If we search long enough for a primate individual with the perfect genotype, we just might find it. This effort of bioprospecting involves genotyping populations of primate species, or any other species relevant to a particular virus, at loci that mediate virus infection. My search for permissive *CD4* alleles in small populations of primates has already turned up positive, and if this search is expanded to include more individuals and more species, even more desirable alleles may be revealed. The beauty of this technique is that these naturally occurring alleles can be mixed and matched using standardized breeding techniques, and do not require any sort of genetic therapy.

The success of bioprospecting lies in an understanding of the critical host-virus interactions that take place during any given virus infection. Numerous genetic loci can be tested for HIV-1 because this is the most studied virus in the history of virology, and there are probably even more critical loci that we have yet to discover. To generalize

bioprospecting for other viruses, it will be necessary to generate a detailed map of host-virus interactions and understand how these interactions mediate infection in a candidate model organism.

Modularity of TRIMs: how a protein family is evolving before our eyes

The tripartite motif containing (TRIM) family of proteins contains over 70 members with diverse functions (304-309). All TRIM proteins have a core domain composed of a RING E3 ligase, one or two B-Boxes that mediate protein-protein interactions, and a coiled-coil domain that is critical for multimerization and functionality of TRIMs (265). The C-terminus of TRIM proteins is variable, although many contain a B30.2 domain, including TRIM25 and TRIM5α. This protein family is unique to metazoans and is highly variable in almost every genera in which it has been examined (16, 135, 310-312). It is even copy number variable within humans, such that certain individuals actually encode over 100 TRIM genes (135).

Perhaps the most interesting case of a TRIM gene is the TRIM5-CypA fusion. This gene has been independently generated in several primate species and is the consequence of a LINE1-mediated insertion of the *CypA* open reading frame into the C-terminus of the *TRIM5* locus (196, 197, 235, 236). TRIM5-CypA is one of the most potent restriction factors against HIV-1, and its mechanism highlights a beautiful evolutionary trick. HIV-1 actually interacts with monomeric CypA to perform its life cycle, and under normal conditions CypA promotes HIV-1 replication. However, when TRIM5-CypA interacts with HIV-1, the core domain of TRIM5 inhibits virus replication

(244). Thus, the fusion of an antiviral TRIM5 with a protein needed by HIV-1, CypA, generated a highly potent restriction factor that functions by tricking HIV-1 into recruiting monomeric CypA. This experiment conducted by nature highlights the modular and flexible nature of TRIMs. And so there are pieces of TRIMs that are involved in recognition of viruses, and other pieces that are involved in the actual restriction activity, whether that be degradation of the target or inhibition of a critical process, such as transcription of influenza vRNPs in the case of TRIM25 (see Chapter 6).

In my future studies I would like to further characterize the modularity of TRIM proteins. Although nature has given us an excellent example in TRIM5-CypA, I believe that this is only the beginning. Further molecular tinkering could reveal more highly potent restriction factors. By mixing and matching domains from TRIM proteins, I hope to generate a molecule with broad specificity against many different virus families. Already, in Chapter 6, I modified the antiretroviral TRIM5α protein to be restrictive against influenza A virus. And, critically, the domain that mediates antiretroviral activity is distinct from that which mediates restriction against influenza A virus. This is further evidence of TRIM modularity, and it also shows that increased specificity can be achieved. There are a handful of other examples where TRIM domains have been swapped, mainly to achieve better expression for purification purposes, or to study interactions with lentiviral capsids (101, 158, 209, 245, 256, 258, 260).

Using the TRIM5-CypA fusion example as a guide, I will take advantage of the growing body of host-virus interactions to generate synthetic TRIM fusions that harbor pieces of proteins that a given virus needs for its life cycle. There will be physical

considerations that I will have to take into account, namely, the size of the target molecule and the mode of inhibition for a particular TRIM protein. These rules will need to be better characterized and, in doing so, I will build a molecular TRIM toolkit that can be used to build an antiviral TRIM protein from scratch.

Conclusions

In this era of high-throughput data and genomics, a challenge we face as scientists is to incorporate large amounts of data from divergent fields into hypothesis driven research. The goal of these efforts is typically to discover novel functions of proteins, or to understand which genes mediate a particular process or disease. However, the use of evolution as our guide can provide a shortcut, whereby signatures of positive selection can highlight regions of a protein that are mediating a critical process, such as viral infection. An added benefit is that we learn how these regions have changed over time, which provides a list of mutations that allow us to fine tune function.

The most exciting part of this field is that it is still brand new. There is no telling where the field of viral molecular evolution might take us. Studying viruses has always led to important biological discoveries, whether they be esoteric or medically relevant. And now, as we strengthen our evolutionary lens, ancient tales of viruses past are sharply coming into focus, bringing with them the secret tools to fight the viruses of today.

References

1. Haldane JBS (1949) Disease and Evolution. *La ricercha scientifica* 19:68–76.

2. Johnson WE (2010) Endless forms most viral. *PLoS Genet* 6:e1001210.

3. Lander ES et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

4. Trowsdale J, Parham P (2004) Mini-review: defense strategies and immunity-related genes. *Eur J Immunol* 34:7–17.

5. Emerman M, Malik HS (2010) Paleovirology--modern consequences of ancient viruses. *PLoS Biol* 8:e1000301.

6. Dawkins R, Krebs JR (1979) Arms races between and within species. *Proc R Soc Lond B Biol Sci* 205:489–511.

7. Worobey M, Bjork A, Wertheim J (2007) Point, Counterpoint: The Evolution of Pathogenic Viruses and their Human Hosts. *Annu. Rev. Ecol. Evol. Syst.*:1–29.

8. Liu S-Y, Sanchez DJ, Cheng G (2011) New developments in the induction and antiviral effectors of type I interferon. *Curr Opin Immunol* 23:57–64.

9. Parham P et al. (2010) Primate-specific regulation of natural killer cells. *J. Med. Primatol.* 39:194–212.

10. Bowie AG, Unterholzner L (2008) Viral evasion and subversion of pattern-recognition receptor signalling. *Nat Rev Immunol* 8:911–922.

11. Blanco-Melo D, Venkatesh S, Bieniasz PD (2012) Intrinsic cellular defenses against human immunodeficiency viruses. *Immunity* 37:399–411.

12. Goila-Gaur R, Strebel K (2008) HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology* 5:51.

13. Evans DT, Serra-Moreno R, Singh RK, Guatelli JC (2010) BST-2/tetherin: a new component of the innate immune response to enveloped viruses. *Trends Microbiol.* 18:388–396.

14. Nisole S, Stoye JP, Saïb A (2005) TRIM family proteins: retroviral restriction and antiviral defence. *Nat. Rev. Microbiol.* 3:799–808.

15. Emerman M (2006) How TRIM5alpha defends against retroviral invasions. *Proc Natl Acad Sci USA* 103:5249–5250.

16. Johnson WE, Sawyer SL (2009) Molecular evolution of the antiretroviral TRIM5 gene. *Immunogenetics* 61:163–176.

17. Sawyer SL, Wu LI, Emerman M, Malik HS (2005) Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc Natl Acad Sci USA* 102:2832–2837.

18. Gupta RK et al. (2009) Mutation of a single residue renders human tetherin resistant to HIV-1 Vpu-mediated depletion. *PLoS Pathog* 5:e1000443.

19. Lim ES, Malik HS, Emerman M (2010) Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. *J Virol* 84:7124–7134.

20. McNatt MW et al. (2009) Species-Specific Activity of HIV-1 Vpu and Positive Selection of Tetherin Transmembrane Domain Variants. *PLoS Pathog* 5:e1000300.

21. O'Brien SJ, Moore JP (2000) The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. *Immunol. Rev.* 177:99–111.

22. Bushman FD et al. (2009) Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog* 5:e1000437.

23. Watanabe T, Watanabe S, Kawaoka Y (2010) Cellular networks involved in the influenza virus life cycle. *Cell Host and Microbe* 7:427–439.

24. Demogines A et al. (2010) Ancient and recent adaptive evolution of primate non-homologous end joining genes. *PLoS Genet* 6:e1001169.

25. Demogines A, Farzan M, Sawyer SL (2012) Evidence for ACE2-Utilizing Coronaviruses (CoVs) Related to Severe Acute Respiratory Syndrome CoV in Bats. *J Virol* 86:6350–6353.

26. Demogines A, Abraham J, Choe H, Farzan M, Sawyer SL (2013) Dual host-virus arms races shape an essential housekeeping protein. *PLoS Biol* 11:e1001571.

27. Meyerson NR et al. (2014) Positive selection of primate genes that promote HIV-1 replication. *Virology* 454-455:291–298.

28. Kirmaier A et al. (2010) TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new species. *PLoS Biol* 8.

29. Wain LV et al. (2007) Adaptation of HIV-1 to its human host. *Mol Biol Evol* 24:1853–1860.

30. Naffakh N, Tomoiu A, Rameix-Welti M-A, van der Werf S (2008) Host restriction of avian influenza viruses at the level of the ribonucleoproteins. *Annu. Rev. Microbiol.* 62:403–424.

31. van Valen L (1973) A new evolutionary law. *Evolutionary Theory* 1:1–30.

32. Chen Z et al. (1998) Natural infection of a homozygous delta24 CCR5 red-capped mangabey with an R2b-tropic simian immunodeficiency virus. *J Exp Med* 188:2057–2065.

33. Riddick NE et al. (2010) A novel CCR5 mutation common in sooty mangabeys reveals SIVsmm infection of CCR5-null natural hosts and efficient alternative coreceptor use in vivo. *PLoS Pathog* 6:e1001064.

34. Mosier DE (2009) How HIV changes its tropism: evolution and adaptation? *Curr Opin HIV AIDS* 4:125–130.

35. Schliekelman P, Garner C, Slatkin M (2001) Natural selection and resistance to HIV. *Nature* 411:545–546.

36. Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet* 32:569–577.

37. O'Brien SJ, Nelson GW (2004) Human genes that limit AIDS. *Nat Genet* 36:565–574.

38. Sawyer SL, Emerman M, Malik HS (2004) Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol* 2:E275.

39. Brass AL et al. (2009) The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* 139:1243–1254.

40. O'Connor SL et al. (2010) MHC heterozygote advantage in simian

immunodeficiency virus-infected Mauritian cynomolgus macaques. *Sci Transl Med* 2:22ra18.

41. Carrington M et al. (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 283:1748–1752.

42. Newman RM et al. (2006) Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5. *Proceedings of the National Academy of Sciences* 103:19134–19139.

43. and Analysis Consortium TCS (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87.

44. Tareen SU, Emerman M (2011) Human Trim5α has additional activities that are uncoupled from retroviral capsid recognition. *Virology* 409:113–120.

45. Pertel T et al. (2011) TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* 472:361–365.

46. Cao W et al. (2009) Regulation of TLR7/9 responses in plasmacytoid dendritic cells by BST2 and ILT7 receptor interaction. *Journal of Experimental Medicine* 206:1603–1614.

47. Malim MH (2009) APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philos Trans R Soc Lond, B, Biol Sci* 364:675–687.

48. Kosiol C et al. (2008) Patterns of positive selection in six Mammalian genomes. *PLoS Genet* 4:e1000144.

49. Rhesus Macaque Genome Sequencing and Analysis Consortium et al. (2007) Evolutionary and Biomedical Insights from the Rhesus Macaque Genome. *Science* 316:222–234.

50. Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167–170.

51. Elde NC, Child SJ, Geballe AP, Malik HS (2009) Protein kinase R reveals an evolutionary model for defeating viral mimicry. *Nature* 457:485–489.

52. Perron MJ, Stremlau M, Sodroski J (2006) Two surface-exposed elements of the B30.2/SPRY domain as potency determinants of N-tropic murine leukemia virus restriction by human TRIM5alpha. *J Virol* 80:5631–5636.

53. Kaiser SM, Malik HS, Emerman M (2007) Restriction of an extinct

retrovirus by the human TRIM5alpha antiviral protein. *Science* 316:1756–1758.

54. Yap MW, Nisole S, Stoye JP (2005) A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. *Curr Biol* 15:73–78.

55. Stremlau M, Perron M, Welikala S, Sodroski J (2005) Species-specific variation in the B30.2(SPRY) domain of TRIM5alpha determines the potency of human immunodeficiency virus restriction. *J Virol* 79:3139–3145.

56. Li Y, Li X, Stremlau M, Lee M, Sodroski J (2006) Removal of arginine 332 allows human TRIM5alpha to bind human immunodeficiency virus capsids and to restrict infection. *J Virol* 80:6738–6744.

57. Keckesova Z, Ylinen LMJ, Towers GJ (2004) The human and African green monkey TRIM5alpha genes encode Ref1 and Lv1 retroviral restriction factor activities. *Proc Natl Acad Sci USA* 101:10780–10785.

58. Perron MJ et al. (2004) TRIM5alpha mediates the postentry block to N-tropic murine leukemia viruses in human cells. *Proc Natl Acad Sci USA* 101:11827–11832.

59. Yap MW, Nisole S, Lynch C, Stoye JP (2004) Trim5alpha protein restricts both HIV-1 and murine leukemia virus. *Proc Natl Acad Sci USA* 101:10786–10791.

60. Hatziioannou T, Perez-Caballero D, Yang A, Cowan S, Bieniasz PD (2004) Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5alpha. *Proc Natl Acad Sci USA* 101:10774–10779.

61. Xu H et al. (2004) A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. *Proc Natl Acad Sci USA* 101:5652–5657.

62. Bogerd HP, Doehle BP, Wiegand HL, Cullen BR (2004) A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. *Proc Natl Acad Sci USA* 101:3770–3774.

63. Schröfelbauer B, Chen D, Landau NR (2004) A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). *Proc Natl Acad Sci USA* 101:3927–3932.

64. Simon V et al. (2005) Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog* 1:e6.

65. Sauter D et al. (2009) Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. *Cell Host Microbe* 6:409–421.

66. Palacios E et al. (1998) Parallel evolution of CCR5-null phenotypes in humans and in a natural host of simian immunodeficiency virus. *Curr Biol* 8:943–946.

67. Wilson SJ et al. (2008) Rhesus macaque TRIM5 alleles have divergent antiretroviral specificities. *J Virol* 82:7243–7247.

68. Song B et al. (2005) The B30.2(SPRY) domain of the retroviral restriction factor TRIM5alpha exhibits lineage-specific length and sequence variation in primates. *J Virol* 79:6111–6121.

69. Ohkura S, Yap MW, Sheldon T, Stoye JP (2006) All three variable regions of the TRIM5alpha B30.2 domain can contribute to the specificity of retrovirus restriction. *J Virol* 80:8554–8565.

70. Delport W, Scheffler K, Seoighe C (2008) Frequent toggling between alternative amino acids is driven by selection in HIV-1. *PLoS Pathog* 4:e1000242.

71. Reszka N, Zhou C, Song B, Sodroski JG, Knipe DM (2010) Simian TRIM5alpha proteins reduce replication of herpes simplex virus. *Virology* 398:243–250.

72. Kuo R-L, Zhao C, Malur M, Krug RM (2010) Influenza A virus strains that circulate in humans differ in the ability of their NS1 proteins to block the activation of IRF3 and interferon-β transcription. *Virology* 408:146–158.

73. Lu J, Pan Q, Rong L, Liu S-L, Liang C (2011) The IFITM Proteins Inhibit HIV-1 Infection. *J Virol* 85:2126–2137.

74. Huang I-C et al. (2011) Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. *PLoS Pathog* 7:e1001258.

75. Kerns JA, Emerman M, Malik HS (2008) Positive Selection and Increased Antiviral Activity Associated with the PARP-Containing Isoform of Human

Zinc-Finger Antiviral Protein. *PLoS Genet* 4:e21.

76. Telenti A, Johnson WE (2012) Host Genes Important to HIV Replication and Evolution. *Cold Spring Harb Perspect Med* 2:a007203.

77. Bloom JD, Gong LI, Baltimore D (2010) Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* 328:1272–1275.

78. Bush RM, Bender CA, Subbarao K, Cox NJ, Fitch WM (1999) Predicting the evolution of human influenza A. *Science* 286:1921–1925.

79. Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet.* 47:97–120.

80. Nielsen R (2005) Molecular signatures of natural selection. *Annu. Rev. Genet.* 39:197–218.

81. Kelley JL, Swanson WJ (2008) Positive Selection in the Human Genome: From Genome Scans to Biological Significance. *Annu. Rev. Genom. Human Genet.* 9:143–160.

82. Hurst LD (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet* 18:486.

83. Duggal NK, Malik HS, Emerman M (2011) The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. *J Virol* 85:11361–11371.

84. Lim ES et al. (2012) The Ability of Primate Lentiviruses to Degrade the Monocyte Restriction Factor SAMHD1 Preceded the Birth of the Viral Accessory Protein Vpx. *Cell Host and Microbe* 11:194–204.

85. Laguette N et al. (2012) Evolutionary and Functional Analyses of the Interaction between the Myeloid Restriction Factor SAMHD1 and the Lentiviral Vpx Protein. *Cell Host and Microbe* 11:205–217.

86. Compton AA, Emerman M (2013) Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses. *PLoS Pathog* 9:e1003135.

87. Mitchell PS et al. (2012) Evolution-Guided Identification of Antiviral Specificity Determinants in the Broadly Acting Interferon-Induced Innate Immunity Factor MxA. *Cell Host and Microbe* 12:598–604.

88. Patel MR, Loo Y-M, Horner SM, Gale M, Malik HS (2012) Convergent evolution of escape from hepaciviral antagonism in primates. *PLoS Biol* 10:e1001282.

89. Kaelber JT et al. (2012) Evolutionary Reconstructions of the Transferrin Receptor of Caniforms Supports Canine Parvovirus Being a Re-emerged and Not a Novel Pathogen in Dogs. *PLoS Pathog* 8:e1002666.

90. Demogines A, Truong KA, Sawyer SL (2012) Species-Specific Features of DARC, the Primate Receptor for Plasmodium vivax and Plasmodium knowlesi. *Mol Biol Evol* 29:445–449.

91. Demogines A, Abraham J, Choe H, Farzan M, Sawyer SL (2013) Dual host-virus arms races shape an essential housekeeping protein. *PLoS Biol* 11:e1001571.

92. Demogines A et al. (2010) Ancient and recent adaptive evolution of primate non-homologous end joining genes. *PLoS Genet* 6:e1001169.

93. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556.

94. Yang Z (2005) Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Mol Biol Evol* 22:1107–1118.

95. Locke DP et al. (2011) Comparative and demographic analysis of orang-utan genomes. *Nature* 469:529–533.

96. Scally A et al. (2012) Insights into hominid evolution from the gorilla genome sequence. *Nature* 483:169–175.

97. Clark AG (2003) Inferring Nonneutral Evolution from Human-Chimp-Mouse Orthologous Gene Trios. *Science* 302:1960–1963.

98. Nielsen R et al. (2005) A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees. *PLoS Biol* 3:e170.

99. Ortiz M et al. (2009) Evolutionary trajectories of primate genes involved in HIV pathogenesis. *Mol Biol Evol* 26:2865–2875.

100. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.

101. Schaller T et al. (2011) HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency.

PLoS Pathog 7:e1002439.

102. Bieniasz PD (2012) An overview of intracellular interactions between immunodeficiency viruses and their hosts. *AIDS* 26:1243–1254.

103. Friedrich BM et al. (2011) Host factors mediating HIV-1 replication. *Virus Research* 161:101–114.

104. Fadel HJ, Poeschla EM (2011) Retroviral restriction and dependency factors in primates and carnivores. *Vet. Immunol. Immunopathol.* 143:179–189.

105. Goff SP (2007) Host factors exploited by retroviruses. *Nat. Rev. Microbiol.* 5:253–263.

106. Friedel CC, Haas J (2011) Virusâ€"host interactomes and global models of virus-infected cells. *Trends Microbiol.* 19:501–508.

107. Hsu T-H, Spindler KR (2012) Identifying Host Factors That Regulate Viral Infection. *PLoS Pathog* 8:e1002772.

108. Shaw ML (2011) The host interactome of influenza virus presents new potential targets for antiviral drugs. *Rev. Med. Virol.* 21:358–369.

109. Houzet L, Jeang KT (2011) Genome-wide screening using RNA interference to study host factors in viral replication and pathogenesis. *Experimental Biology and Medicine* 236:962–967.

110. Hirsch AJ (2010) The use of RNAi-based screens to identify host proteins involved in viral replication. *Future Microbiology* 5:303–311.

111. Fernandez-Garcia M-D, Mazzon M, Jacobs M, Amara A (2009) Pathogenesis of Flavivirus Infections: Using and Abusing the Host Cell. *Cell Host and Microbe* 5:318–328.

112. Ptak RG et al. (2008) Short Communication:Cataloguing the HIV Type 1 Human Protein Interaction Network. *AIDS Research and Human Retroviruses* 24:1497–1502.

113. Fu W et al. (2009) Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Research* 37:D417–D422.

114. Goff SP (2008) Knockdown screens to knockout HIV-1. *Cell* 135:417–420.

115. Brass AL et al. (2008) Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 319:921–926.

116. König R et al. (2008) Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* 135:49–60.

117. Yeung ML, Houzet L, Yedavalli VSRK, Jeang K-T (2009) A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. *J Biol Chem* 284:19463–19473.

118. Zhou H et al. (2008) Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* 4:495–504.

119. Jäger S et al. (2012) Global landscape of HIV-human protein complexes. *Nature* 481:365–370.

120. Bushman FD et al. (2009) Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog* 5:e1000437.

121. de Chassey B, Meyniel-Schicklin L, Aublin-Gex A, André P, Lotteau V (2012) Genetic screens for the control of influenza virus replication: from meta-analysis to drug discovery. *Mol. BioSyst.* 8:1297.

122. Compton AA, Hirsch VM, Emerman M (2012) The Host Restriction Factor APOBEC3G and Retroviral Vif Protein Coevolve due to Ongoing Genetic Conflict. *Cell Host and Microbe* 11:91–98.

123. OhAinle M, Kerns JA, Malik HS, Emerman M (2006) Adaptive Evolution and Antiviral Activity of the Conserved Mammalian Cytidine Deaminase APOBEC3H. *J Virol* 80:3853–3862.

124. Stoye JP (2012) Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat. Rev. Microbiol.*

125. Gifford RJ (2012) Viral evolution in deep time: lentiviruses and mammals. *Trends in Genetics* 28:89–100.

126. Daugherty MD, Malik HS (2012) Rules of engagement: molecular insights from host-virus arms races. *Annu. Rev. Genet.* 46:677–700.

127. Meyerson NR, Sawyer SL (2011) Two-stepping through time: mammals and viruses. *Trends Microbiol.* 19:286–294.

128. Duggal NK, Emerman M (2012) Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol* 12:687–695.

129. Wertheim JO, Worobey M (2007) A challenge to the ancient origin of

SIVagm based on African green monkey mitochondrial genomes. *PLoS Pathog* 3:e95.

130. Malim MH, Bieniasz PD (2012) HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harb Perspect Med* 2:a006940–a006940.

131. Malim MH, Emerman M (2008) HIV-1 Accessory Proteins—Ensuring Viral Survival in a Hostile Environment. *Cell Host and Microbe* 3:388–398.

132. McCarthy, KR *et al*. (2013) Gain-of-sensitivity mutations in a TRIM5-resistant primary isolate of pathogenic SIV identify two independent conserved determinants of TRIMalpha specificity. *PLoS Pathog* 9:e1003352

133. Sawyer SL, Elde NC (2012) A cross-species view on viruses. *Current Opinion in Virology* 2:561–568.

134. Tareen SU, Sawyer SL, Malik HS, Emerman M (2009) An expanded clade of rodent Trim5 genes. *Virology* 385:473–483.

135. Han K, Lou DI, Sawyer SL (2011) Identification of a Genomic Reservoior for New *TRIM* Genes in Primate Genomes. *PLoS Genet* 7.

136. Sawyer SL, Emerman M, Malik HS (2007) Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. *PLoS Pathog* 3:e197.

137. Demogines A, Abraham J, Choe H, Farzan M, Sawyer SL (2013) Dual Host-Virus Arms Races Shape an Essential Housekeeping Protein. *PLoS Biol* 11:e1001571.

138. Sawyer SL, Malik HS (2006) Positive selection of yeast nonhomologous end-joining genes and a retrotransposon conflict hypothesis. *Proceedings of the National Academy of Sciences* 103:17614–17619.

139. Zhang ZD, Weinstock G, Gerstein M (2008) Rapid evolution by positive Darwinian selection in T-cell antigen CD4 in primates. *J Mol Evol* 66:446–456.

140. Parmley JL, Hurst LD (2007) How Common Are Intragene Windows with K A > K S Owing to Purifying Selection on Synonymous Mutations? *J Mol Evol* 64:646–655.

141. Comeron JM (1999) K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics* 15:763–764.

142. Yamashita M, Emerman M (2004) Capsid Is a Dominant Determinant of Retrovirus Infectivity in Nondividing Cells. *J Virol* 78:5670–5678.

143. Abramoff MD, Magalhaes PJ, Ram SJ (2004) Image Processing with ImageJ. *Biophotonics International*.

144. Barraza RA, Poeschla EM (2008) Human gene therapy vectors derived from feline lentiviruses. *Vet. Immunol. Immunopathol.* 123:23–31.

145. Kwong PD et al. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393:648–659.

146. Schmid K, Yang Z (2008) The trouble with sliding windows and the selective pressure in BRCA1. *PLoS One* 3:e3746.

147. Yang Z, Bielawski J (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol (Amst)* 15:496–503.

148. Wilen CB, Tilton JC, Doms RW (2012) HIV: Cell Binding and Entry. *Cold Spring Harb Perspect Med* 2:a006866–a006866.

149. Strambio-De-Castillia C, Niepel M, Rout MP (2010) The nuclear pore complex: bridging nuclear transport and gene regulation. *Nat Rev Mol Cell Biol* 11:490–501.

150. Di Nunzio F et al. (2012) Human nucleoporins promote HIV-1 docking at the nuclear pore, nuclear import and integration. *PLoS One* 7:e46037.

151. Matreyek KA, Engelman A (2011) The Requirement for Nucleoporin NUP153 during Human Immunodeficiency Virus Type 1 Infection Is Determined by the Viral Capsid. *J Virol* 85:7818–7827.

152. Woodward CL, Prakobwanakit S, Mosessian S, Chow SA (2009) Integrase Interacts with Nucleoporin NUP153 To Mediate the Nuclear Import of Human Immunodeficiency Virus Type 1. *J Virol* 83:6522–6533.

153. Gallo DE, Hope TJ (2012) Knockdown of MAP4 and DNAL1 produces a post-fusion and pre-nuclear translocation impairment in HIV-1 replication. *Virology* 422:13–21.

154. Woodard AH et al. (2011) NY-BR-1 and PAX8 Immunoreactivity in Breast, Gynecologic Tract, and Other CK7+ Carcinomas: Potential Use for Determining Site of Origin. *American Journal of Clinical Pathology* 136:428–435.

155. Hombrouck A et al. (2007) Virus Evolution Reveals an Exclusive Role for LEDGF/p75 in Chromosomal Tethering of HIV. *PLoS Pathog* 3:e47.

156. Lee K et al. (2010) Flexible Use of Nuclear Import Pathways by HIV-1. *Cell Host and Microbe* 7:221–233.

157. Jäger D et al. (2007) NY-BR-1 is a differentiation antigen of the mammary gland. *Applied immunohistochemistry & molecular morphology*. 15:77–83.

158. Lee K et al. (2012) HIV-1 Capsid-Targeting Domain of Cleavage and Polyadenylation Specificity Factor 6. *J Virol* 86:3851–3860.

159. Hopkins AL, Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* 1:727–730.

160. OhAinle M, Kerns JA, Li MMH, Malik HS, Emerman M (2008) Antiretroelement Activity of APOBEC3H Was Lost Twice in Recent Human Evolution. *Cell Host and Microbe* 4:249–259.

161. McDougle RM, Hultquist JF, Stabell AC, Sawyer SL, Harris RS (2013) D316 is critical for the enzymatic activity and HIV-1 restriction potential of human and rhesus APOBEC3B. *Virology* 441:31–39.

162. Jäger S et al. (2012) Global landscape of HIV-human protein complexes. *Nature* 481:365–370.

163. Whitehead KA, Dahlman JE, Langer RS, Anderson DG (2011) Silencing or Stimulation? siRNA Delivery and the Immune System. *Annu. Rev. Chem. Biomol. Eng.* 2:77–96.

164. Park H et al. (2010) Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. *Nat Genet* 42:400–405.

165. Bushman FD et al. (2013) Bringing it all together. *AIDS* 27:835–838.

166. Fellay J, Shianna KV, Telenti A, Goldstein DB (2010) Host Genetics and HIV-1: The Final Phase? *PLoS Pathog* 6:e1001033.

167. Telenti A, Goldstein DB (2006) Genomics meets HIV-1. *Nat. Rev. Microbiol.* 4:865–873.

168. Vallender EJ (2004) Positive selection on the human genome. *Human Molecular Genetics* 13:R245–R254.

169. Cherepanov P (2004) Identification of an Evolutionarily Conserved Domain in Human Lens Epithelium-derived Growth Factor/Transcriptional Co-activator p75 (LEDGF/p75) That Binds HIV-1 Integrase. *Journal of Biological Chemistry* 279:48883–48892.

170. Valle-Casuso JC et al. (2012) TNPO3 Is Required for HIV-1 Replication after Nuclear Import but prior to Integration and Binds the HIV-1 Core. *J Virol* 86:5931–5936.

171. De Iaco A et al. (2013) TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm. *Retrovirology* 10:20.

172. Hatziioannou T, Evans DT (2012) Animal models for HIV/AIDS research. *Nat. Rev. Microbiol.* 10:852–867.

173. Harris RS, Hultquist JF, Evans DT (2012) The restriction factors of human immunodeficiency virus. *Journal of Biological Chemistry* 287:40875–40883.

174. Humes D, Emery S, Laws E, Overbaugh J (2012) A species-specific amino acid difference in the macaque CD4 receptor restricts replication by global circulating HIV-1 variants representing viruses from recent infection. *J Virol* 86:12472–12483.

175. Fomsgaard A et al. (1995) Receptor Function of CD4 Structures from African Green Monkey and Pig-Tail Macaque for Simian Immunodeficiency Virus, SIVsm, SIVagm, and Human Immunodeficiency Virus Type-1. *Viral Immunology* 8.

176. Wong SK et al. (2009) A New World primate deficient in tetherin-mediated restriction of human immunodeficiency virus type 1. *J Virol* 83:8771–8780.

177. Hatziioannou T et al. (2014) HIV-1-induced AIDS in monkeys. *Science* 344:1401–1405.

178. Karlsson GB et al. (1997) Characterization of molecularly cloned simian-human immunodeficiency viruses causing rapid CD4+ lymphocyte depletion in rhesus monkeys. *J Virol* 71:4218–4225.

179. Humes D, Overbaugh J (2011) Adaptation of subtype a human immunodeficiency virus type 1 envelope to pig-tailed macaque cells. *J Virol* 85:4409–4420.

180. Puffer BA et al. (2002) CD4 Independence of Simian Immunodeficiency Virus Envs Is Associated with Macrophage Tropism, Neutralization

Sensitivity, and Attenuated Pathogenicity. *J Virol* 76:2595–2605.

181. Long EM, Rainwater SMJ, Lavreys L, Mandaliya K, Overbaugh J (2002) HIV type 1 variants transmitted to women in Kenya require the CCR5 coreceptor for entry, regardless of the genetic complexity of the infecting virus. *AIDS Research and Human Retroviruses* 18:567–576.

182. Blish CA et al. (2009) Cross-subtype neutralization sensitivity despite monoclonal antibody resistance among early subtype A, C, and D envelope variants of human immunodeficiency virus type 1. *J Virol* 83:7783–7788.

183. Li M et al. (2005) Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* 79:10108–10125.

184. Li M et al. (2006) Genetic and neutralization properties of subtype C human immunodeficiency virus type 1 molecular env clones from acute and early heterosexually acquired infections in Southern Africa. *J Virol* 80:11776–11790.

185. Mirzabekov T et al. (1999) Enhanced expression, native purification, and characterization of CCR5, a principal HIV-1 coreceptor. *J Biol Chem* 274:28745–28750.

186. Pacheco B, Basmaciogullari S, LaBonte JA, Xiang S-H, Sodroski J (2008) Adaptation of the human immunodeficiency virus type 1 envelope glycoproteins to new world monkey receptors. *J Virol* 82:346–357.

187. Fomsgaard A et al. (1997) Relation between phylogeny of African green monkey CD4 genes and their respective simian immunodeficiency virus genes. *J. Med. Primatol.* 26:120–128.

188. Hvilsom C et al. (2008) Genetic subspecies diversity of the chimpanzee CD4 virus-receptor gene. *Genomics* 92:322–328.

189. LaBonte JA, Babcock GJ, Patel T, Sodroski J (2002) Blockade of HIV-1 infection of New World monkey cells occurs primarily at the stage of virus entry. *J Exp Med* 196:431–445.

190. White TE et al. (2014) Effects of human SAMHD1 polymorphisms on HIV-1 susceptibility. *Virology* 460-461:34–44.

191. Schaner P et al. (2001) Episodic evolution of pyrin in primates: human mutations recapitulate ancestral amino acid states. *Nat Genet* 27:318–321.

192. Wang JH et al. (2001) Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. *Proc Natl Acad Sci USA* 98:10799–10804.

193. Huang CC et al. (2004) Structural basis of tyrosine sulfation and VH-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120. *Proceedings of the National Academy of Sciences* 101:2706–2711.

194. Perelman P et al. (2011) A molecular phylogeny of living primates. *PLoS Genet* 7:e1001342.

195. Sharp PM, Hahn BH (2011) Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med* 1:a006841.

196. Sayah DM, Sokolskaja E, Berthoux L, Luban J (2004) Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* 430:569–573.

197. Nisole S, Lynch C, Stoye JP, Yap MW (2004) A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. *Proc Natl Acad Sci USA* 101:13324–13328.

198. Ribeiro IP et al. (2005) Evolution of cyclophilin A and TRIMCyp retrotransposition in New World primates. *J Virol* 79:14998–15003.

199. Krupp A et al. (2013) APOBEC3G polymorphism as a selective barrier to cross-species transmission and emergence of pathogenic SIV and AIDS in a primate host. *PLoS Pathog* 9:e1003641.

200. Sawyer SL, Wu LI, Akey JM, Emerman M, Malik HS (2006) High-Frequency Persistence of an Impaired Allele of the Retroviral Defense Gene TRIM5α in Humans. *Current Biology* 16:95–100.

201. Duggal NK, Fu W, Akey JM, Emerman M. (2013) Identification and antiviral activity of common polymorphisms in the APOBEC3 locus in human populations. *Virology* 443:329-337

202. Chang JJ et al. (2012) Differential regulation of toll-like receptor pathways in acute and chronic HIV-1 infection. *AIDS* 26:533–541.

203. Mothe BR et al. (2003) Expression of the Major Histocompatibility Complex Class I Molecule Mamu-A*01 Is Associated with Control of Simian Immunodeficiency Virus SIVmac239 Replication. *J Virol* 77:2736–2740.

204. Yant LJ et al. (2006) The High-Frequency Major Histocompatibility

Complex Class I Allele Mamu-B*17 Is Associated with Control of Simian Immunodeficiency Virus SIVmac239 Replication. *J Virol* 80:5074–5077.

205. Loffredo JT et al. (2007) Mamu-B*08-Positive Macaques Control Simian Immunodeficiency Virus Replication. *J Virol* 81:8827–8832.

206. Ward JM, Vallender EJ (2012) The resurgence and genetic implications of New World primates in biomedical research. *Trends Genet* 28:586–591.

207. Hofmann W et al. (1999) Species-specific, postentry barriers to primate immunodeficiency virus infection. *J Virol* 73:10020–10028.

209. Diaz-Griffero F et al. (2006) Requirements for capsid-binding and an effector function in TRIMCyp-mediated restriction of HIV-1. *Virology* 351:404–419.

210. Yamashita M, Emerman M (2006) Retroviral infection of non-dividing cells: Old and new perspectives. *Virology* 344:88–93.

211. Szymborska A et al. (2013) Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* 341:655–658.

212. D'Angelo MA, Hetzer MW (2008) Structure, dynamics and function of nuclear pore complexes. *Trends Cell Biol*. 18:456–466.

214. Di Nunzio F (2013) New insights in the role of nucleoporins: a bridge leading to concerted steps from HIV-1 nuclear entry until integration. *Virus Res* 178:187–196.

215. Di Nunzio F et al. (2013) Nup153 and Nup98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. *Virology* 440:8–18.

216. Meehan AM et al. (2014) A cyclophilin homology domain-independent role for Nup358 in HIV-1 infection. *PLoS Pathog* 10:e1003969.

217. Yokoyama N et al. (1995) A giant nucleopore protein that binds Ran/TC4. *Nature* 376:184–188.

218. Wu J, Matunis MJ, Kraemer D, Blobel G, Coutavas E (1995) Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. *J Biol Chem* 270:14209–14213.

219. Rabut G, Doye V, Ellenberg J (2004) Mapping the dynamic organization of the nuclear pore complex inside single living cells. *Nat. Cell Biol*. 6:1114–

1121.

220. Bernad R, van der Velde H, Fornerod M, Pickersgill H (2004) Nup358/RanBP2 attaches to the nuclear pore complex via association with Nup88 and Nup214/CAN and plays a supporting role in CRM1-mediated nuclear protein export. *Mol Cell Biol* 24:2373–2384.

221. Hutten S, Wälde S, Spillner C, Hauber J, Kehlenbach RH (2009) The nuclear pore component Nup358 promotes transportin-dependent nuclear import. *J. Cell. Sci.* 122:1100–1110.

222. Zhang R, Mehla R, Chauhan A (2010) Perturbation of host nuclear membrane component RanBP2 impairs the nuclear import of human immunodeficiency virus -1 preintegration complex (DNA). *PLoS One* 5:e15620.

223. Ocwieja KE et al. (2011) HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS Pathog* 7:e1001313.

224. Mamede JI, Sitbon M, Battini J-L, Courgnaud V (2013) Heterogeneous susceptibility of circulating SIV isolate capsids to HIV-interacting factors. *Retrovirology* 10:77.

225. Bichel K et al. (2013) HIV-1 capsid undergoes coupled binding and isomerization by the nuclear pore protein NUP358. *Retrovirology* 10:81.

226. Lin T-Y, Emerman M (2006) Cyclophilin A interacts with diverse lentiviral capsids. *Retrovirology* 3:70.

227. Qi M, Yang R, Aiken C (2008) Cyclophilin A-dependent restriction of human immunodeficiency virus type 1 capsid mutants for infection of nondividing cells. *J Virol* 82:12001–12008.

228. Braaten D, Franke EK, Luban J (1996) Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. *J Virol* 70:3551–3560.

229. Braaten D, Franke EK, Luban J (1996) Cyclophilin A is required for the replication of group M human immunodeficiency virus type 1 (HIV-1) and simian immunodeficiency virus SIV(CPZ)GAB but not group O HIV-1 or other primate immunodeficiency viruses. *J Virol* 70:4220–4227.

230. Li Y, Kar AK, Sodroski J (2009) Target cell type-dependent modulation of

human immunodeficiency virus type 1 capsid disassembly by cyclophilin A. *J Virol* 83:10951–10962.

231. Franke EK, Yuan HE, Luban J (1994) Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* 372:359–362.

232. Homme MB, Carter C, Scarlata S (2005) The Cysteine Residues of HIV-1 Capsid Regulate Oligomerization and Cyclophilin A-Induced Changes. *Biophysical Journal* 88:2078–2088.

233. Fricke T, Brandariz-Nuñez A, Wang X, Smith AB, Diaz-Griffero F (2013) Human Cytosolic Extracts Stabilize the HIV-1 Core. *J Virol* 87:10587–10597.

234. Gamble TR et al. (1996) Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* 87:1285–1294.

235. Malfavon-Borja R, Wu LI, Emerman M, Malik HS (2013) Birth, decay, and reconstruction of an ancient TRIMCyp gene fusion in primate genomes. *Proceedings of the National Academy of Sciences* 110:E583–92.

236. Wilson SJ et al. (2008) Independent evolution of an antiviral TRIMCyp in rhesus macaques. *Proceedings of the National Academy of Sciences* 105:3557–3562.

237. McCarthy KR et al. (2013) Gain-of-sensitivity mutations in a Trim5-resistant primary isolate of pathogenic SIV identify two independent conserved determinants of Trim5α specificity. *PLoS Pathog* 9:e1003352.

238. Wu F et al. (2013) TRIM5 alpha Drives SIVsmm Evolution in Rhesus Macaques. *PLoS Pathog* 9:e1003577.

239. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556.

240. Price AJ et al. (2009) Active site remodeling switches HIV specificity of antiretroviral TRIMCyp. *Nat. Struct. Mol. Biol.* 16:1036–1042.

241. Caines MEC et al. (2012) Diverse HIV viruses are targeted by a conformationally dynamic antiviral. *Nat. Struct. Mol. Biol.* 19:411–416.

242. Ylinen LMJ et al. (2010) Conformational adaptation of Asian macaque TRIMCyp directs lineage specific antiviral activity. *PLoS Pathog* 6:e1001062.

244. Yap MW, Dodding MP, Stoye JP (2006) Trim-cyclophilin A fusion proteins can restrict human immunodeficiency virus type 1 infection at two distinct phases in the viral life cycle. *J Virol* 80:4061–4067.

245. Chan E et al. (2012) Lentiviral gene therapy against human immunodeficiency virus type 1, using a novel human TRIM21-cyclophilin A restriction factor. *Hum. Gene Ther.* 23:1176–1185.

246. Takehisa J et al. (2009) Origin and biology of simian immunodeficiency virus in wild-living western gorillas. *J Virol* 83:1635–1648.

247. Klatt NR, Silvestri G, Hirsch V (2012) Nonpathogenic simian immunodeficiency virus infections. *Cold Spring Harb Perspect Med* 2:a007153.

248. Hahn BH, Shaw GM, De Cock KM, Sharp PM (2000) AIDS as a zoonosis: scientific and public health implications. *Science* 287:607–614.

249. Plantier J-C et al. (2009) A new human immunodeficiency virus derived from gorillas. *Nat Med* 15:871–872.

250. van Rensburg EJ et al. (1998) Simian immunodeficiency viruses (SIVs) from eastern and southern Africa: detection of a SIVagm variant from a chacma baboon. *J Gen Virol* 79 (Pt 7):1809–1814.

251. Jin MJ et al. (1994) Infection of a yellow baboon with simian immunodeficiency virus from African green monkeys: evidence for cross-species transmission in the wild. *J Virol* 68:8454–8460.

252. Bailes E et al. (2003) Hybrid origin of SIV in chimpanzees. *Science* 300:1713.

253. Neel C et al. (2010) Molecular epidemiology of simian immunodeficiency virus infection in wild-living gorillas. *J Virol* 84:1464–1476.

254. Korber B et al. (2000) Timing the ancestor of the HIV-1 pandemic strains. *Science* 288:1789–1796.

255. Yarbrough ML, Mata MA, Sakthivel R, Fontoura BMA (2014) Viral subversion of nucleocytoplasmic trafficking. *Traffic* 15:127–140.

256. Matreyek KA, Yücel SS, Li X, Engelman A (2013) Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. *PLoS Pathog* 9:e1003693.

257. Martin C, Buckler-White A, Wollenberg K, Kozak CA (2013) The avian XPR1 gammaretrovirus receptor is under positive selection and is disabled in bird species in contact with virus-infected wild mice. *J Virol* 87:10094–10104.

258. Liu Z et al. (2013) The interferon-inducible MxB protein inhibits HIV-1 infection. *Cell Host Microbe* 14:398–410.

259. Ozato K, Shin D-M, Chang T-H, Morse HC (2008) TRIM family proteins and their emerging roles in innate immunity. *Nat Rev Immunol* 8:849–860.

260. Ganser-Pornillos BK et al. (2011) Hexagonal assembly of a restricting TRIM5{alpha} protein. *Proc Natl Acad Sci USA* 108:534–539.

261. Wang Q, Tao YJ (2010) *Influenza* (Horizon Scientific Press).

262. Ortiz JR et al. (2014) The burden of influenza-associated critical illness hospitalizations*. *Crit. Care Med.* 42:2325–2332.

263. Kochs G, Garcia-Sastre A, Martinez-Sobrido L (2007) Multiple anti-interferon actions of the influenza A virus NS1 protein. *J Virol* 81:7011–7021.

264. Gack MU et al. (2009) Influenza A virus NS1 targets the ubiquitin ligase TRIM25 to evade recognition by the host viral RNA sensor RIG-I. *Cell Host Microbe* 5:439–449.

265. Sardiello M, Cairo S, Fontanella B, Ballabio A, Meroni G (2008) Genomic analysis of the TRIM family reveals two groups of genes with distinct evolutionary properties. *BMC Evol. Biol.* 8:225.

266. Gack MU et al. (2007) TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* 446:916–920.

267. Zeng W et al. (2010) Reconstitution of the RIG-I Pathway Reveals a Signaling Role of Unanchored Polyubiquitin Chains in Innate Immunity. *Cell* 141:315–330.

268. Rajsbaum R et al. (2012) Species-specific inhibition of RIG-I ubiquitination and IFN induction by the influenza A virus NS1 protein. *PLoS Pathog* 8:e1003059.

269. Kwon SC et al. (2013) The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* 20:1122–1130.

270. Malfavon-Borja R, Sawyer SL, Wu LI, Emerman M, Malik HS (2013) An evolutionary screen highlights canonical and noncanonical candidate antiviral genes within the primate TRIM gene family. *Genome Biol Evol* 5:2141–2154.

271. Elde NC et al. (2012) Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell* 150:831–841.

272. Kawakami E et al. (2011) Strand-specific real-time RT-PCR for distinguishing influenza vRNA, cRNA, and mRNA. *J. Virol. Methods* 173:1–6.

273. Arranz R et al. (2012) The structure of native influenza virion ribonucleoproteins. *Science* 338:1634–1637.

274. Taylor RT et al. (2011) TRIM79α, an Interferon-Stimulated Gene Product, Restricts Tick-Borne Encephalitis Virus Replication by Degrading the Viral RNA Polymerase. *Cell Host and Microbe* 10:185–196.

275. Di Pietro A et al. (2013) TRIM22 inhibits influenza A virus infection by targeting the viral nucleoprotein for degradation. *J Virol* 87:4523–4533.

276. Stremlau M et al. (2004) The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* 427:848–853.

277. Pacheco B, Finzi A, Stremlau M, Sodroski J (2010) Adaptation of HIV-1 to cells expressing rhesus monkey TRIM5α. *Virology* 408:204–212.

278. O'Brien TC, Tauraso NM (1973) Antibodies to type A influenza viruses in sera from nonhuman primates. *Arch Gesamte Virusforsch* 40:359–365.

279. Karlsson EA et al. (2012) Influenza virus infection in nonhuman primates. *Emerging Infect. Dis.* 18:1672–1675.

280. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457.

281. Vreede FT, Brownlee GG (2007) Influenza virion-derived viral ribonucleoproteins synthesize both mRNA and cRNA in vitro. *J Virol* 81:2196–2204.

282. Patel MR, Emerman M, Malik HS (2011) Paleovirologyâ€"ghosts and gifts of viruses past. *Current Opinion in Virology*:1–6.

283. Lavialle C et al. (2013) Paleovirology of “syncytins,” retroviral env genes exapted for a role in placentation. *Philos Trans R Soc Lond, B, Biol Sci* 368:20120507.

284. Rethwilm A, Bodem J (2013) Evolution of foamy viruses: the most ancient of all retroviruses. *Viruses* 5:2349–2374.

285. Han G-Z, Worobey M (2014) A primitive endogenous lentivirus in a colugo: insights into the early evolution of lentiviruses. *Mol Biol Evol*.

286. Stremlau M et al. (2004) The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* 427:848–853.

287. Busnadiego I et al. (2014) Host and viral determinants of Mx2 antiretroviral activity. *J Virol* 88:7738–7752.

288. Xiao H, Killip MJ, Staeheli P, Randall RE, Jackson D (2013) The human interferon-induced MxA protein inhibits early stages of influenza A virus infection by retaining the incoming viral genome in the cytoplasm. *J Virol* 87:13053–13058.

289. Chen G, Liu C-H, Zhou L, Krug RM (2014) Cellular DDX21 RNA helicase inhibits influenza A virus replication but is counteracted by the viral NS1 protein. *Cell Host Microbe* 15:484–493.

291. Meliopoulos VA et al. (2012) Host gene targets for novel influenza therapies elucidated by high-throughput RNA interference screens. *FASEB J*. 26:1372–1386.

292. Fusco DN et al. (2013) A genetic screen identifies interferon-α effector genes required to suppress hepatitis C virus replication. *Gastroenterology* 144:1438–49, 1449.e1–9.

293. Lee AS-Y, Burdeinick-Kerr R, Whelan SPJ (2014) A genome-wide small interfering RNA screen identifies host factors required for vesicular stomatitis virus infection. *J Virol* 88:8355–8360.

294. Choe H, Jemielity S, Abraham J, Radoshitzky SR, Farzan M (2011) Transferrin receptor 1 in the zoonosis and pathogenesis of New World hemorrhagic fever arenaviruses. *Curr. Opin. Microbiol*. 14:476–482.

295. Peng L et al. (2014) Widespread sequence variations in VAMP1 across vertebrates suggest a potential selective pressure from botulinum neurotoxins. *PLoS Pathog* 10:e1004177.

296. Dyer MD, Murali TM, Sobral BW (2008) The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog* 4:e32.

297. Masson P et al. (2014) An integrated ontology resource to explore and study host-virus relationships. *PLoS One* 9:e108075.

298. Haus T et al. (2014) Genome typing of nonhuman primate models: implications for biomedical research. *Trends Genet*:1–6.

299. Vallender EJ, Miller GM (2013) Nonhuman primate models in the genomic era: a paradigm shift. *ILAR J* 54:154–165.

300. Phillips KA et al. (2014) Why primate models matter. *Am J Primatol* 76:801–827.

301. Liu Y et al. (2013) Genetic variants in IL1A and IL1B contribute to the susceptibility to 2009 pandemic H1N1 influenza A virus. *BMC Immunol*. 14:37.

302. Lee MN et al. (2014) Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* 343:1246980.

303. Hatziioannou T et al. (2009) A macaque model of HIV-1 infection. *Proceedings of the National Academy of Sciences* 106:4425–4429.

304. Petrera F, Meroni G (2012) TRIM proteins in development. *Adv. Exp. Med. Biol*. 770:131–141.

305. Liu Y et al. (2014) TRIM3, a tumor suppressor linked to regulation of p21(Waf1/Cip1.). *Oncogene* 33:308–315.

306. Quintás-Cardama A et al. (2014) Loss of TRIM62 Expression Is an Independent Adverse Prognostic Factor in Acute Myeloid Leukemia. *Clin Lymphoma Myeloma Leuk*.

307. Uchil PD et al. (2014) TRIM15 is a focal adhesion protein that regulates focal adhesion disassembly. *J. Cell. Sci*. 127:3928–3942.

308. Tomar D, Singh R (2014) TRIM family proteins; Emerging class of RING E3 ligases as regulator of NF-κB pathway. *Biol. Cell*.

309. Mandell MA et al. (2014) TRIM proteins regulate autophagy and can target autophagic substrates by direct recognition. *Dev. Cell* 30:394–409.

310. Marín I (2012) Origin and diversification of TRIM ubiquitin ligases. *PLoS*

One 7:e50030.

311. Si Z et al. (2006) Evolution of a cytoplasmic tripartite motif (TRIM) protein in cows that restricts retroviral infection. *Proc Natl Acad Sci USA* 103:7454–7459.

312. van der Aa LM et al. (2009) A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. *BMC Biol*. 7:7.

Vita

Nicholas Ryan Meyerson was born in Houston, Texas to Mr. Jeffrey Meyerson and Mrs. Youvella De La Garza Meyerson in 1985. He grew up 25 miles Northeast of Houston in a small (at the time) suburb called Humble, along with his two older siblings, Lisa and Donald. He attended Pine Forest Elementary School, Atasocita Middle School, and Humble High School during grades K-12. In 2003, Nicholas began undergraduate studies at The University of Texas at Austin. He spent the academic year from 2005-2006 studying physics abroad at Queen Mary University in London. He accepted an internship at NASA's Jet Propulsion Laboratory during the Fall 2007 semester. In Spring of 2008 he earned a Bachelor of Science in Physics with High Honors from The University of Texas at Austin. That following summer, he began working towards a Ph.D. in Cell and Molecular Biology at The University of Texas at Austin in The Institute for Cellular and Molecular Biology. He has published two scientific articles as a first author and one as a secondary author; with three more first author manuscripts in preparation and a secondary authorship currently in review. Starting in January of 2015, Nicholas will extend his dissertation work with Dr. Sara Sawyer as a postdoctoral fellow at The University of Colorado Boulder.

Permanent email address: nrmeyerson@gmail.com

This dissertation was typed by Nicholas Ryan Meyerson